2 Digital Waveguide Modeling

The simulation technique used in this work is called digital waveguide modeling. This methodology has recently gained much interest in musical acoustics and computer music, since it is suitable for developing physical models of musical instruments that can be efficiently implemented. This chapter reviews the principles of waveguide modeling and discusses related approaches.

The aim in digital waveguide modeling is to design a discrete-time model that behaves like a physical system. The theory of digital waveguide modeling has been primarily developed by Smith (Smith, 1987, 1992b, 1993a, 1995a). This technique is especially well suited to simulation of one-dimensional resonators, such as a vibrating string, a narrow acoustic tube, or a thin bar. The method was, however, first applied to artificial reverberation using delay line networks (Smith, 1985) and only after that to the synthesis of wind and string instruments (Smith, 1986).

A vibrating string is one of the traditional topics of research in musical acoustics, and waveguide modeling has been applied to this case. In the following we first examine the foundations of waveguide modeling using the string as an example. The motivation for this is that the waveguide string models are nowadays highly developed: the synthesis and also analysis algorithms are studied. The simplest model for a vibrating string—the Karplus–Strong algorithm—and its extensions are considered. A large part of this chapter concentrates on waveguide modeling of acoustic tubes which are the main application area of this work. The final sections discuss relations of waveguide modeling to wave digital filters and to filter structures that are used in digital reverberation.

2.1 Waveguide Modeling of Strings

The most important single idea in waveguide modeling is that approximate traveling wave solutions of wave equations encountered in musical instruments can be used for time-domain simulation.

2.1.1 Wave Equation and the Traveling Wave Solution

The following exposition has been adapted from Smith (1992b). The one-dimensional wave equation for an ideal vibrating string is given by

$$K\frac{\partial^2 y}{\partial t^2} = \varepsilon\frac{\partial^2 y}{\partial x^2} \tag{2.1}$$

where t is time, x is distance, $y(t, x)$ is the string displacement, K is the string tension, and ε is the mass density of the string. The general solution of the wave equation was published by d'Alembert in 1747 and is given by

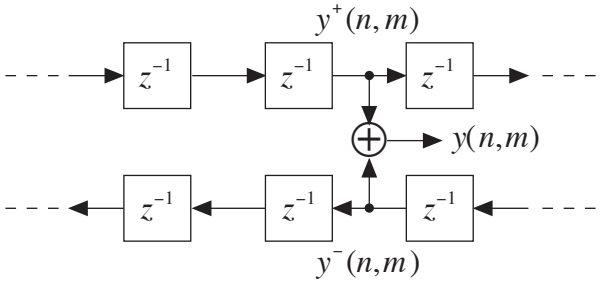

Fig. 2.1 A digital waveguide consists of two delay lines where the signal components travel in opposite directions. The amplitude of the signal at any point (n, m) is obtained summing the two components.

$$y(x,t) = y^+(x - ct) + y^-(x + ct) \tag{2.2}$$

where $y^+(x - ct)$ and $y^-(x + ct)$ are two arbitrary twice-differentiable functions that represent waves traveling in opposite directions with speed $c = \sqrt{K / \varepsilon}$.

When the traveling waveforms are bandlimited lowpass signals, it is straightforward to sample the traveling wave solution of Eq. (2.2). This yields

$$y(n,m) = y^+(n,m) + y^-(n,m) \tag{2.3}$$

Here we have discretized both the temporal variable t and the spatial variable x by substituting

$$\begin{aligned} t &= nT \quad \text{for } n = 0,1,2,\ldots \\ x &= mX \quad \text{for } m = 0,1,2,\ldots \end{aligned} \tag{2.4}$$

where T is the sampling interval, i.e., the inverse of the sampling rate, and X is the spatial sampling interval which is related to T as

$$X = cT \tag{2.5}$$

where c is the speed of sound ($c \approx 340$ m/s).

Equation (2.3) can be interpreted as a *bidirectional delay line* where a fixed sampled waveform $y^+(n,m)$ travels in the positive direction and $y^-(n,m)$ in the negative direction, as illustrated in Fig. 2.1. The amplitude of the signal at any point is obtained by summing the values of the components at that point, as indicated by Eq. (2.3). This structure is called a *digital waveguide* (see Fig. 2.1). Usually the spatial index m is not used since it depends on the temporal index n.

2.1.2 Incorporating Losses and Dispersion

Above we have considered modeling of a lossless ideal string. This kind of a string is purely hypothetical but it has several properties in common with a realistic string. Smith (1990) has pointed out that the wave equation including frequency-dependent losses and dispersion can also be solved analytically. This results in a modified traveling wave solution that can be sampled and used for waveguide simulation.

Resistive losses (i.e., yielding terminations of the string, drag of air, and internal friction of the material) add to the wave equation a term that is proportional to $\partial y / \partial t$,

i.e., the time-derivative of displacement. Smith (1990) has shown that this kind of distributed losses can be modeled by a real gain factor associated with each unit delay of the waveguide. The multipliers between each input and output point can be commuted and combined thus preserving the computational efficiency of the waveguide system.

Frequency-dependent losses, such as those due to the bridge where the strings are attached, can be modeled by placing a linear digital filter between each unit delay element of the waveguide. These elements may also be commuted and combined so that between each input-output pair the total loss is modeled correctly. The computational load of the commuted system may be two or three orders of magnitude less than that of the straightforward one.

Physical systems are always *passive* and *stable*. In waveguide simulation it is of paramount importance to preserve the passive nature of the real-world system. A waveguide model includes a feedback loop, and if the loop gain exceeds unity at any frequency, the system becomes unstable. Hence, digital filters to be included in a digital waveguide have to be designed so that their magnitude response is equal to or less than unity at all frequencies. This perspective has to be remembered when any linear system—such as a fractional delay filter—is added to a waveguide system.

2.1.3 Karplus–Strong Algorithm and Its Extensions

Karplus and Strong (1983) proposed a simple algorithm for the synthesis of plucked string sounds. The method is based on the idea that a wavetable, i.e., a table containing a sampled waveform of an audio signal, is modified while it is read. This technique was found (Smith, 1983) to be a special case of a string simulation studied by McIntyre, Woodhouse, and Schumacher (McIntyre and Woodhouse, 1979; McIntyre *et al.*, 1983), and it was extended with these ideas in mind by Jaffe and Smith (1983). The Karplus–Strong (KS) algorithm is an important predecessor of current waveguide models because it established the usefulness of highly simplified cases and recently it has led to quite detailed string instruments models (see, e.g., Karjalainen *et al.*, 1993; Karjalainen and Välimäki, 1993; Smith, 1993; Välimäki *et al.*, 1995a).

The block diagram of the KS algorithm is shown in Fig. 2.2. The system is a recursive comb filter. The delay line (or wavetable) is initialized with white noise. The output of the delay line is fed to a lowpass filter that is called the *loop filter*. The filtering result is the output of the system and it is also fed back to the delay line.

This technique does not produce a purely repetitive output signal as does the basic wavetable synthesis (sampling) technique. Namely, those frequency components of the random signal that coincide with the resonances of the comb filter will attenuate slower than the other components. Thus, the signal in the delay line progressively turns into a pseudo-periodic signal with a clearly perceivable fundamental frequency.

In the original KS model, the loop filter $H_1(z)$ is a two-tap averager which is easy to implement without multiplications. This simple filter, however, cannot match the frequency-dependent damping of a physical string. We feel that an FIR filter is not suited to this purpose since its order should be rather high to match the desired characteristics.

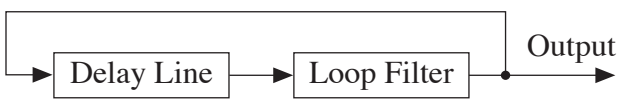

Fig. 2.2 The Karplus–Strong algorithm.

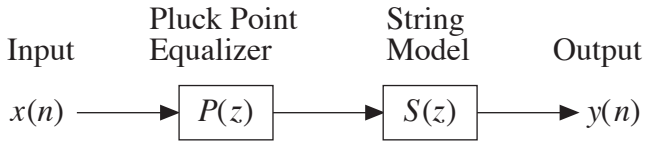

Fig. 2.3 The generalized Karplus–Strong model consists of a string model $S(z)$ cascaded with a plucking point equalizer $P(z)$.

Instead we suggest the use of an IIR lowpass filter for simulating the damping characteristics of a physical string. From a physical point of view, it seems well motivated to use a second or third-order all-pole filter as a loop filter (Fletcher and Rossing, 1991). It is meaningful to use the simplest satisfying loop filter since one such filter is needed for the model of each string. Also, the loop filter is not a static filter but its coefficients should be changed as a function of the string length and other playing parameters. This is easiest to achieve using a low-order filter.

Furthermore, it is important to remember that the loop filter's magnitude response must not exceed unity at any frequency since otherwise the system becomes unstable. When the magnitude response of the filter is less than unity at all frequencies the resulting sound will gradually attenuate. The timbre will also vary over time if the magnitude response is not flat, because the harmonics of the signal will attenuate at different rates. Dispersion can be brought about using a filter with a nonlinear phase response.

We have found that a reasonable approximation for the frequency-dependent damping in a string may be obtained using a first-order all-pole filter

$$H_1(z) = g\frac{1 + a_1}{1 + a_1 z^{-1}} \tag{2.6}$$

where g is the gain of the filter at $\omega = 0$, and a_1 is the filter coefficient that determines the cutoff frequency of the filter. For $H_1(z)$ to be a *stable lowpass* transfer function, we require that $-1 < a_1 < 0$. The numerator $1 + a_1$ scales the frequency response (divided by g) at $\omega = 0$ to unity thus allowing control of the low-frequency gain using the coefficient g. We require that $|g| \leq 1$.

A comb filter $P(z)$ may be cascaded with the string model to cause the filtering effect due to the plucking position (Jaffe and Smith, 1983). Its transfer function is

$$P(z) = 1 + z^{-M} R_f(z) \tag{2.7}$$

where $R_f(z)$ is the reflection filter of one end of the string model and M is the delay (in samples) between the excitation point in the upper and lower line of the original waveguide model. In practice $R_f(z)$ in (2.7) need not be modeled very accurately since its magnitude response is only slightly less than unity at all frequencies. In practice the most remarkable difference between the direct and reflected excitation is the inverted phase and thus the filter can be replaced by a constant multiplier $r_f = -1 + \varepsilon$ where ε is a small nonnegative real number. This modified system that we call a *generalized KS model* is illustrated in Fig. 2.3.

One of the major problems of the basic Karplus–Strong algorithm is that the fundamental frequency of the tone cannot be accurately controlled. The fundamental frequency f_0 of the synthetic signal is determined by the length of the delay line, that is

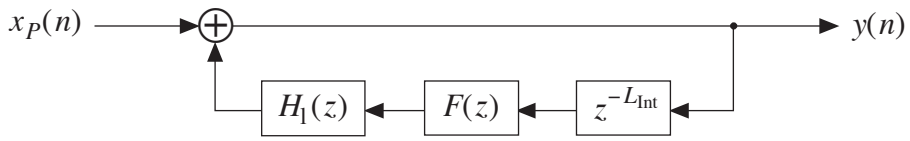

Fig. 2.4 The block diagram of the generalized Karplus–Strong model $S(z)$ for a vibrating string. The filter $F(z)$ is an FIR or allpass-type fractional delay filter, L_{Int} is the length of the delay line (in samples), and $H_1(z)$ is the loop filter that accounts for losses and dispersion. The input signal $x_P(n)$ has been filtered with the pluck point equalizer (see, Fig. 2.3).

$$f_0 = \frac{f_s}{M + 1/2} \tag{2.8}$$

where f_s is the sampling rate, M is the length of the delay line in samples, and 1/2 refers to the phase delay (or, equivalently, the group delay) of the two-point averaging filter. It is seen that Eq. (2.8) is a function of the integer-valued variable M. Thus the fundamental frequency is quantized and an arbitrary pitch cannot be produced.

The length of the delay line loop, and consequently the fundamental frequency f_0 of the synthetic signal, can be accurately controlled by adding a *fractional delay filter* in the feedback loop of the Karplus–Strong algorithm. The block diagram of the string model $S(z)$ is shown in Fig. 2.4. Jaffe and Smith (1983) introduced a first-order allpass filter for producing the desired fractional delay. Later, a linear interpolator (Sullivan, 1990) and a third-order Lagrange interpolator (Karjalainen and Laine, 1991) have been used for this task. Chapter 3 of this work discusses various digital filter approximations for the fractional delay.

2.1.4 Calibration of the Waveguide String Model

The synthesis system includes three parts that completely determine the character of the synthetic sound: the string model, the plucking point equalizer, and the input sequence. This implies that to calibrate the model to some particular instrument it is needed to estimate the values for the length of the delay line L, the coefficients of the loop filter $H_1(z)$, the delay M and the parameter r_f of the plucking-point equalizer, and the input signal $x(n)$. In this section, parameter estimation procedures are described that will extract these values.

The delay length L (in samples) determines the fundamental frequency f_0 of the synthetic signal as follows

$$f_0 = \frac{f_s}{L} \tag{2.9}$$

where f_s is the sample rate. For pitch detection we have used a well-known method based on the windowed autocorrelation function (Rabiner, 1977). An estimate for the pitch is obtained by searching for the maximum of this function for each frame. Median filtering may be used to remove erroneous results from the pitch estimate sequence. We have successfully used three-point median filtering for this purpose.

The pitch of a plucked string tone is usually a function of time, decreasing slightly after the attack and reaching a steady value within about 0.5 s. The problem then is to

determine the best estimate $\hat{f}_0$ for the fundamental frequency in a perceptual sense. In practice a good solution is to use the average pitch value after some 500 ms as the nominal value, since it is important to have a reliable pitch estimate towards the end of the note. This improves the quality of synthesized tones. When the estimate $\hat{f}_0$ for the fundamental frequency has been chosen, the effective length L of the delay line can be computed as

$$L = \frac{f_s}{\hat{f}_0} \tag{2.10}$$

When also the loop filter has been designed, we can subtract its phase delay $\tau_H(\omega)$ (at the estimated fundamental frequency) from L and define the nominal fractional delay L_f as

$$L_f = L - \tau_H(\hat{\omega}_0) - \lfloor L - \tau_H(\hat{\omega}_0) \rfloor \tag{2.11}$$

where $\hat{\omega}_0 = 2\pi\hat{f}_0$, and $\lfloor \cdot \rfloor$ is the floor function [see Eq. (3.12)]. The delay L_f is used as the desired delay in the design of the fractional delay filter $F(z)$. In practice the phase delay of this filter can be expressed as $\tau_f(\hat{\omega}_0) = M_f + L_f$, where $M_f \in \mathbb{Z}_+$ is the integral part of the phase delay that depends on the order of the fractional delay filter and L_f is the fractional part.

Next we discuss how to measure the damping of the string as a function of frequency†. All losses including the reflection at the two ends of the string are incorporated in the loop filter $H_l(z)$. In order to design a loop filter we need to estimate the damping factors at the harmonic frequencies of the sound signal. This is achieved using the *short-time Fourier transform* (STFT) and tracking the amplitude of each harmonic.

The STFT of signal $y(n)$ is a sequence of discrete Fourier transforms (DFT) defined as

$$Y_m(k) = \sum_{n=0}^{N-1} w(n)y(n + mH)e^{-j\omega_k n} \quad \text{for } m = 0,\ 1,\ 2,\ldots \tag{2.12a}$$

with

$$\omega_k = \frac{2\pi k}{N} \quad \text{for } k = 0,\ 1,\ 2,\ldots,N-1 \tag{2.12b}$$

where N is the length of the DFT, $w(n)$ is a window function, and H is the hop size or time advance (in samples) per frame. In practice we compute each DFT using the FFT algorithm. To obtain a suitable compromise between time and frequency resolution, we use a window length of four times the period length of the signal. The overlap of the windows is 50% implying that H is 0.5 times the window length. We apply excessive *zero-padding* by filling the signal buffer with zeros to reach $N = 4096$ in order to increase the resolution in the frequency domain. The spectral peaks corresponding to harmonics can be found from the magnitude spectrum by first finding the nearest local minimum at both sides of an assumed maximum. The largest magnitude between these

† Smith (1993b) has independently developed a very similar analysis method for the waveguide string model.

two local minima is assumed to be the harmonic peak. The estimates for the frequency and magnitude of the peak are fine-tuned by applying *parabolic interpolation* around the local maximum and solving for the value and location of the maximum of this interpolating function. For details, see Smith and Serra (1987) or Serra (1989). The number of harmonics to be detected is typically $N_\mathrm{h} < 20$ (for the acoustic guitar). The STFT analysis is applied to the portion of the signal that starts before the attack and ends after some 0.5–1 s after the attack.

The spectral peak detection results in a sequence of magnitude and frequency pairs for each harmonic. The sequence of magnitude values for one harmonic is called the *envelope curve* of that harmonic. A straight line is matched to each envelope curve on a dB scale since ideally the magnitude of every harmonic should decay exponentially, i.e., linearly on a logarithmic scale. Measurements show that this idealized case is rarely met in practice, and many different kinds of deviations are common. It is possible to decrease the error in the fit by starting the fit at the maximum of the envelope curve and by terminating it before the data gets mixed with the noise floor, e.g., after the decay of about 40 dB. As a result, a collection of slopes β_k for $k = 1, 2, \ldots, N_\mathrm{h}$ is obtained. The corresponding *loop gain* of the string model at the harmonic frequencies is computed as

$$G_k = 10^{\frac{\beta_k L}{20H}} \quad \text{for } k = 1,\ 2,\ldots, N_h \tag{2.13}$$

where β_k are the slopes of the envelopes, and H the hop size. The sequence G_k determines the prototype magnitude response of the loop filter $H_1(\omega)$ at the harmonic frequencies kf_0, where $k = 1, 2, \ldots, N_\mathrm{h}$.

The desired phase response for $H_1(\omega)$ can be determined based on the frequency estimates of the harmonics. We do not, however, try to match the phase response of the filter. There are two principal reasons for this. First, we believe that it is much more important to match the time constants of the partials of the synthetic sound than the frequencies of the partials since quite small deviations from harmonic frequencies occur in the case of plucked string instruments. In case we wanted to resynthesize strongly inharmonic tones or a piano with properly stiff strings, the phase response of the loop filter should be considered. The second reason is that as we want to use a low-order loop filter, the complex approximation of the frequency response would not be very successful. Thus we restrict ourselves to magnitude approximation only in the loop filter design.

The human ear is sensitive to the change of decay rate of a sinusoid, and in practice we measure the time constant of some 20 lowest harmonics with the object of matching the frequency response of the loop filter $H_1(\omega)$ to that data. Since we use a first-order loop filter, it is clear that there are more restrictions than unknowns in this problem, and only an approximate solution is possible. We have decided to use weighted least squares design (Välimäki *et al.*, 1995a).

It is reasonable to choose a weighting function $W(G_k)$ that gives a larger weight to the errors in the time constants of the slowly decaying harmonics since the hearing tends to focus on their decay. A candidate for such a weighting function is

$$W(G_k) = \frac{1}{1 - G_k} \tag{2.14}$$

We require that $0 \leq G_k < 1$ for all k, which is a physically reasonable assumption since the system to be modeled is passive and stable.

The gain of the loop filter at low frequencies, i.e., g, can be chosen based on the loop gain values of the lowest harmonics. In many cases it is good enough to set $g = G_0$ (Välimäki *et al.*, 1995a). The value for the coefficient a_1 that minimizes the error function can then be found easily by differentiating the error function with respect to a_1 and finding the zero of the derivative. In practice we find a near-optimal solution in the following way: the value of the derivative is evaluated and, depending on the sign of the result, a_1 is changed by a small increment, the derivative is evaluated again, and so on. After the derivative has reached a very small value, the iteration is terminated and the final value for a_1 is used in the synthesis model.

We have verified the convergence of this design procedure in practice by analyzing signals generated using the synthesis model. The loop filter designed based on the analysis data yields the same filter parameters—within numerical accuracy—as used in the synthesis. Also, matches with natural tones have been found to be satisfactory in many cases (Välimäki *et al.*, 1995a).

It is well understood that when a string is set into to vibration by plucking it, the sound signal will lack those harmonics that have a node at the plucking point (see, e.g., Fletcher and Rossing, 1991). However, in general, the string is not plucked exactly at the node of any of the lowest harmonics, and since the amplitude of the higher harmonics is considerably small anyway, it is not always possible to accurately detect the plucking point by simply searching for the lacking harmonics in the magnitude spectrum. Another practical problem may occur because of nonlinear behavior of the string. Namely, the amplitude of vibration of a weak harmonic can gain energy from other modes so that its amplitude begins to rise reaching a maximum about 100 ms after the attack and then begin to decay (Legge and Fletcher, 1984). This can often be seen in the analysis of harmonic envelopes of the guitar. For these reasons we believe that a more comprehensive understanding of the effect of the plucking point can be achieved by studying the time-domain behavior of the string in terms of the short-time autocorrelation function. A frequency-domain technique for estimating the plucking point has been reported by Bradley *et al.* (1995).

Estimation of the plucking position is an inherently difficult problem since a recorded tone can include contributions of several delays of approximately the same magnitude, e.g., early reflections from objects near the player, such as the floor, the ceiling, or a wall. To minimize the effect of these additional factors, we suggest analysis of tones recorded in an anechoic chamber.

It is, however, not absolutely obligatory to estimate and model the effect of the plucking position. Its contribution can be left in the excitation signal obtained using inverse filtering, which is described in the following section.

2.1.5 Excitation Signal of the Waveguide String Model

The input signal $x(n)$ of the waveguide string model can be estimated using *inverse filtering*, i.e., filtering of the signal $y(n)$—which is now assumed to be the output of the model—with the inverted transfer function $S^{-1}(z)$ of the string model. The transfer function of the string model shown in Fig. 2.4 can be expressed as

$$S(z) = \frac{1}{1 - z^{-L_{\text{Int}}} F(z) H_1(z)} \tag{2.15}$$

where L_{Int} is the length of the delay line (in samples), $F(z)$ is the transfer function of the fractional delay filter, and $H_1(\omega)$ is the transfer function of the loop filter. If $S(z)$ had zeros outside the unit circle, the inverse filter $S^{-1}(z)$ would be unstable. In that case, inverse filtering would not yield an acceptable result. Fortunately, this is not a problem in practice as can be seen by substituting (2.6) into (2.15). Inverting this equation yields

$$S^{-1}(z) = \frac{1 + a_1 z^{-1} - g(1 + a_1 z^{-1}) z^{-L_I} F(z)}{1 + a_1 z^{-1}} \tag{2.16}$$

This technique is simple to apply but since the order of the loop filter is low, the harmonics are not canceled accurately using the inverse filter $S^{-1}(z)$. The resulting $x(n)$ may suffer from high-frequency noise. The low-order loop filter was chosen primarily since it is efficient to implement the synthesis algorithm. However, inverse filtering is an off-line procedure where accuracy is much more important than efficiency. We could thus design a higher order filter to be used in the inverse filtering as suggested by Laroche and Meillier (1994).

Another deficiency of the inverse filtering technique described above is that it does not take into account the nonlinear effects of the physical strings. This could be accounted for by using a time-varying inverse filter to cancel the transfer function of the string. This filter can be designed based on the STFT analysis that was discussed in the previous section.

The residual signals resulting from the analysis and inverse filtering of an original string instrument tone can be directly applied as the excitation to the string model. The excitation signal can be shortened by windowing the first 100 ms of the residual signal using, e.g., the right half of a Hamming window. This truncated signal can be used as the excitation of the synthesis model. In the case of the electric guitar, the length of the residual signal used in resynthesis can be reduced to about 50 ms without any significant loss of information. This is due to the lack of body resonances.

Usually the residual signal of an acoustic guitar tone includes only a few low-frequency components of the body of the instrument. They have large time constants. The most prominent ones are typically the air mode (the Helmholtz resonance) and the lowest mode of the back plate. Välimäki *et al.* (1995a) as well as Smith and Van Duyne (1995) have suggested the extraction of these two strongly resonating modes. Once they have been removed from the residual signal, the resulting signal is very short. The two body resonances can be simulated using second-order all-pole filters.

2.1.6 Discussion

In this section we have discussed waveguide modeling of string instruments. We have seen that these synthesis models have reached an advanced level: the model is compact and efficient to implement but still mainly based on the physics of the instrument. The analysis method enabled resynthesis of string tones or copying the timbre of a given plucked string instrument. Smith and Van Duyne (1995) have used similar waveguide techniques and some of their extensions for modeling the piano (see also Van Duyne and Smith, 1995).

Plucked string instruments can be considered relatively easy from the viewpoint of physical modeling, since their resonator is a simple linear system. Later in this chapter we will tackle modeling of wind instruments which are strongly nonlinear systems with a complicated linear resonator: a bore that contains finger holes or valves. For these reasons modeling is more complicated. Furthermore, nonlinearities make analysis tools difficult to devise.

2.2 Waveguide Modeling of Cylindrical Tubes

The cylindrical tubes that are of interest in speech communication and in musical acoustics are narrow with respect to the wavelength of the sound waves propagating in them. In narrow tubes, only plane waves can propagate at frequencies below a certain critical frequency. Above this frequency, the wavelength is no longer much larger than the diameter of the tube and thus also transverse modes can propagate. The main interest in this work is the plane-wave propagation since it can be simulated using a one-dimensional digital waveguide model.

An articulatory speech synthesizer is a model of the human speech production system that generates speech-like signals based on slowly-varying physiological parameters, such as lung pressure, shape of the tongue, and lip opening. A common feature of articulatory speech synthesizers is that they contain a model for the wave propagation in the human vocal tract. There are, nevertheless, several approaches available to do this.

Kelly and Lochbaum (1962) used a simple time-domain model of the vocal tract. Their method has now become classic and is called the *Kelly–Lochbaum* (KL) *model*. Later several research groups have used versions of the technique (e.g., Kabasawa *et al.*, 1983; Liljencrants, 1985; Meyer *et al.*, 1989). Other approaches include solving of partial differential equations (see, e.g., Maeda, 1982) and a chain-matrix method (Sondhi and Schroeter, 1987). All of the methods are based on the linear wave equation. Since the model is also assumed to be linear, the problem can be handled either in the time or in the frequency domain.

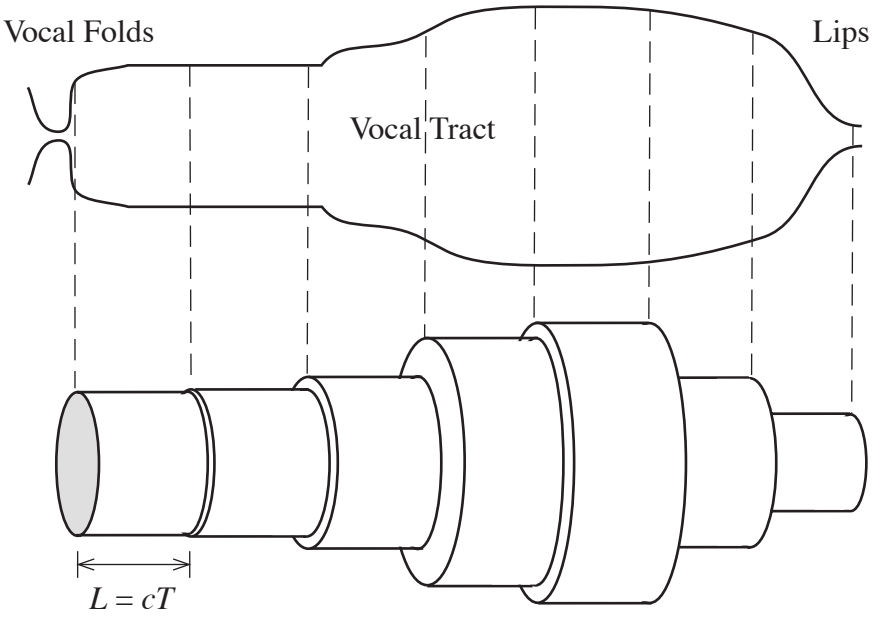

Fig. 2.5 Vocal tract profile and its approximation by cylindrical tube sections.

The Kelly–Lochbaum speech synthesis model is of interest here since it is based on the same idea as the digital waveguide technique. The KL model has not been used for commercial speech synthesizers but rather for research purposes. Recently, Cook (1989, 1990, 1993) has applied the KL model to the synthesis of the singing voice. His model can be used by composers of computer music, and it may also generate stimuli for psychoacoustic testing.

The KL model implements a piecewise zeroth-order polynomial approximation for the continuously variable shape of the tube. This is illustrated in Fig. 2.5. Each tube section of the approximated system has its length L determined by the sampling interval T and the speed of sound c. At the junction of tube sections of different diameter the acoustic wave is scattered, i.e., the wave is partly reflected from the discontinuity and partly transmitted through it. Digital implementation of the KL model is discussed in the following.

2.2.1 The One-Dimensional Wave Equation

The wave equation for a plane wave in a lossless cylindrical tube is a partial differential equation given by

$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{\partial^2 p}{\partial x^2} \tag{2.17}$$

where c is the propagation speed of the sound wave ($\approx$ 340 m/s), p is the longitudinal pressure displacement in the tube, t is time, and x is the distance. This equation may be derived using Newton's second law of motion (Fletcher and Rossing, 1991). The second-order terms have been neglected in Eq. (2.17) and thus it is sometimes called the *linearized wave equation*. This is, of course, an approximation, but it appears to be quite adequate even with high sound pressures. This equation is practically always valid when considering the human vocal tract or bores of musical instruments.

The general solution to the one-dimensional wave equation has the form

$$p(x,t) = p^+(x - ct) + p^-(x + ct) \tag{2.18}$$

where p^+ and p^- are arbitrary continuous functions. The term $p^+(x - ct)$ represents a pressure wave component traveling in the positive x direction at speed c whereas $p^-(x + ct)$ represents a pressure wave traveling in the negative x direction at the same speed. Provided that the signal traveling in the tube is *bandlimited*, this solution can be discretized and a discrete-time waveguide model can be constructed as explained in Section 2.1.1.

The exact shape of the pressure and volume velocity waves in a cylindrical tube depends on the initial conditions and the boundary conditions. Typical cases for boundary conditions are a closed and an open end.

2.2.2 A Model Composed of Coupled Uniform Tube Sections

The characteristic impedance Z of an acoustic tube is defined by

$$Z = \frac{\rho c}{A} = \frac{\rho c}{\pi a^2} \tag{2.19}$$

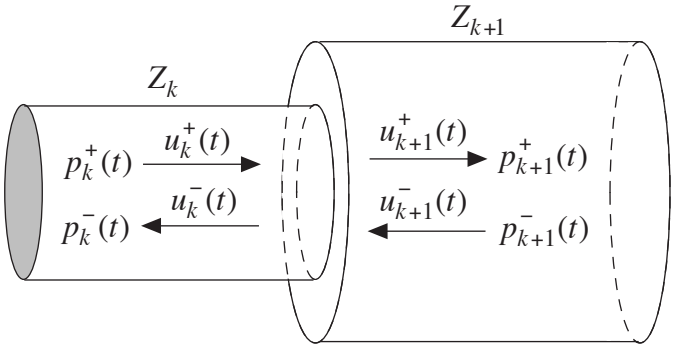

Fig. 2.6 Junction of two uniform tubes of different diameter.

where ρ is the density of air in the tube, c is the speed of sound, and A is the cross-sectional area of the tube the diameter of which is $2a$. The characteristic impedance determines the ratio of pressure and volume velocity components in the tube, that is

$$Z = \frac{p^+(x,t)}{u^+(x,t)} = \frac{p^-(x,t)}{u^-(x,t)} \tag{2.20}$$

This is an analogy of Ohm's law in electrical circuit theory.

A discontinuity in diameter of a cylindrical tube changes the characteristic impedance and, consequently, disturbs the acoustic wave. The incident wave is only partly transmitted through the discontinuity and part of its energy is reflected back. This phenomenon is called *scattering*. Here we shall consider a system consisting of M coupled uniform tubes of different diameter but same length. In Chapter 4, an extended model with variable-length sections is presented.

The computational model for a discontinuity of a cylindrical acoustic tube is derived below for both the volume velocity and the pressure waves. The derivations are based on a classical assumption that pressure and volume velocity are continuous at every junction. Sondhi (1983) has stated that, strictly speaking, the pressure is discontinuous and therefore an additional *acoustic inductance* should be added at every junction. We shall, however, use the classical approach and Sondhi's improvement is left for future work.

2.2.3 KL Scattering Junction for Volume Velocity Waves

The reflection and transmission coefficients needed for modeling the discontinuity of impedance in a cylindrical tube are derived below. Figure 2.6 introduces the variable quantities that are used in the derivation.

The total sound pressure in the kth tube section is expressed as

$$p_k(x,t) = Z_k\left[u_k^+(t-\tau) + u_k^-(t+\tau)\right] \tag{2.21}$$

where $\tau = x/c$ and Z_k is the characteristic impedance of the kth tube section. The total volume velocity in the kth segment is the difference of the two components

$$u_k(x,t) = u_k^+(t-\tau) - u_k^-(t+\tau) \tag{2.22}$$

The pressure has to be continuous at the connection of the tubes and the net flow in

the contiguous tubes has to be the same:

$$\begin{cases} p_k(L,t) = p_{k+1}(0,t) \\ u_k(L,t) = u_{k+1}(0,t) \end{cases} \tag{2.23}$$

where $L = cT$ is the length of tube sections. We continue by substituting Eqs. (2.21) and (2.22) into Eq. (2.23). This yields

$$\begin{cases} Z_k\left[u_k^+(t-\tau) + u_k^-(t+\tau)\right] = Z_{k+1}\left[u_{k+1}^+(t) + u_{k+1}^-(t)\right] \\ u_k^+(t-\tau) - u_k^-(t+\tau) = u_{k+1}^+(t) - u_{k+1}^-(t) \end{cases} \tag{2.24}$$

From this pair of equations, it is easy to derive the equations describing the volume velocity components u_k^- and u_{k+1}^+. They are

$$u_k^-(t+\tau) = \frac{Z_{k+1} - Z_k}{Z_{k+1} + Z_k} u_k^+(t-\tau) + \frac{2Z_{k+1}}{Z_{k+1} + Z_k} u_{k+1}^-(t) \tag{2.25a}$$

$$u_{k+1}^+(t) = \frac{2Z_k}{Z_{k+1} + Z_k} u_k^+(t-\tau) + \frac{Z_k - Z_{k+1}}{Z_{k+1} + Z_k} u_{k+1}^-(t) \tag{2.25b}$$

Based on the above equations it is reasonable to define the *reflection coefficient* for the kth junction in the positive direction as (Liljencrants, 1985, p. 1-3; Fletcher and Rossing, 1991, pp. 147–148)†

$$r_k = \frac{Z_{k+1} - Z_k}{Z_{k+1} + Z_k} = \frac{A_k - A_{k+1}}{A_k + A_{k+1}} \tag{2.26a}$$

where Z_k and A_k denote the acoustic impedance and cross-sectional area of the kth tube section, respectively. The latter form has been devised using relation (2.19). It is seen that $|r_k| \le 1$, since the areas (and consequently the impedances) are always nonnegative. The reflection coefficient of the same junction in the negative direction is opposite in sign. The *transmission coefficient* in the positive direction is written as

$$t_k = \frac{2Z_k}{Z_{k+1} + Z_k} = \frac{2A_{k+1}}{A_k + A_{k+1}} = 1 - r_k \tag{2.26b}$$

The transmission coefficient in the negative direction is $1 + r_k$. Note that if $Z_{k+1} = Z_k$, no scattering occurs.

Equations (2.25a) and (2.25b) may be expressed more compactly using Eq. (2.26a), that is

$$u_k^-(t+\tau) = r_k u_k^+(t-\tau) + (1 + r_k) u_{k+1}^-(t) \tag{2.27a}$$

$$u_{k+1}^+(t) = (1 - r_k) u_k^+(t-\tau) - r_k u_{k+1}^-(t) \tag{2.27b}$$

† There is some confusion in the literature about the definition of the reflection coefficient for the two-port junction. In some references the definition is opposite in sign with respect to the definition used here (see, e.g., Rabiner and Schafer, 1978, pp. 84–85).

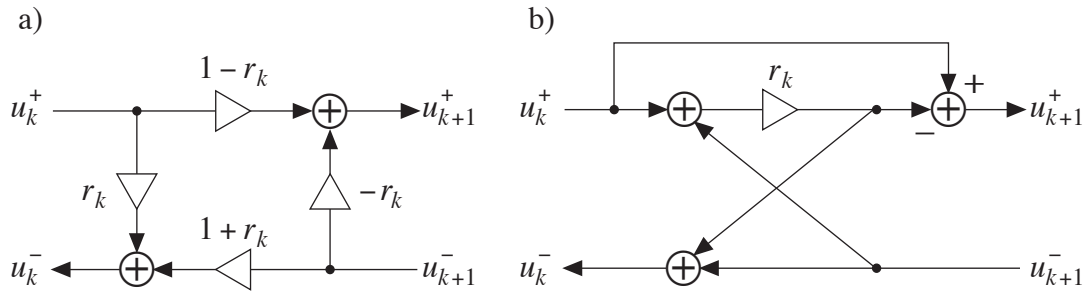

Fig. 2.7 a) The two-port scattering junction for acoustic volume velocity signals and b) an equivalent one-multiplier junction.

The signal flow diagram describing the scattering of volume velocity at a junction of two uniform tube segments is illustrated in Fig. 2.7a. A more economical implementation of this junction is obtained by rewriting equations (2.27a) and (2.27b) in the following way:

$$u_k^-(t+\tau) = u_{k+1}^-(t) + w(t) \tag{2.28a}$$

$$u_{k+1}^+(t) = u_k^+(t-\tau) - w(t) \tag{2.28b}$$

where

$$w(t) = r_k\Big[u_k^+(t-\tau) + u_{k+1}^-(t)\Big] \tag{2.28c}$$

Since the term $w(t)$ is common to Eqs. (2.28a) and (2.28b), it needs to be computed only once. This one-multiplier configuration of the two-port junction is depicted in Fig. 2.7b.

2.2.4 KL Scattering Junction for Pressure Waves

The two-port junction may also be derived for pressure signals. The total pressure in the kth section is given by (2.18) and the net flow (total volume velocity) is written as

$$u_k(x,t) = u_k^+(t-\tau) - u_k^-(t+\tau) = \frac{1}{Z_k}\Big[p_k^+(t-\tau) - p_k^-(t+\tau)\Big] \tag{2.29}$$

The application of continuity requirements for both pressure and volume velocity results in the following pair of equations

$$\begin{cases} \dfrac{1}{Z_k}\Big[p_k^+(t-\tau) - p_k^-(t+\tau)\Big] = \dfrac{1}{Z_{k+1}}\Big[p_{k+1}^+(t-\tau) - p_{k+1}^-(t+\tau)\Big] \\ p_k^+(t-\tau) + p_k^-(t+\tau) = p_{k+1}^+(t) + p_{k+1}^-(t) \end{cases} \tag{2.30}$$

By eliminating the term $p_{k+1}^+(t)$ it is easy to solve for $p_k^-(t+\tau)$ and, similarly by eliminating $p_k^-(t+\tau)$, for $p_{k+1}^+(t)$

$$p_k^-(t+\tau) = \frac{Z_{k+1} - Z_k}{Z_{k+1} + Z_k}\, p_k^+(t-\tau) + \frac{2Z_k}{Z_{k+1} + Z_k}\, p_{k+1}^-(t) \tag{2.31a}$$

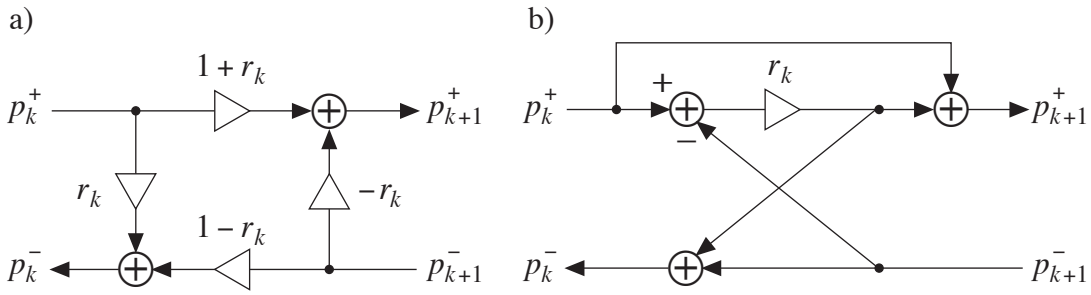

Fig. 2.8 a) The two-port scattering junction for the acoustic pressure signal and b) the equivalent one-multiplier junction.

$$p_{k+1}^+(t) = \frac{2Z_{k+1}}{Z_{k+1} + Z_k} p_k^+(t - \tau) + \frac{Z_k - Z_{k+1}}{Z_{k+1} + Z_k} p_{k+1}^-(t) \tag{2.31b}$$

The same reflection coefficient as in the case of volume velocity signals can still be used. By using Eq. (2.26a), the pair of Eqs. (2.31a) and (2.31b) may be expressed as

$$p_k^-(t + \tau) = r_k p_k^+(t - \tau) + (1 - r_k) p_{k+1}^-(t) \tag{2.32a}$$

$$p_{k+1}^+(t) = (1 + r_k) p_k^+(t - \tau) - r_k p_{k+1}^-(t) \tag{2.32b}$$

The scattering of the pressure wave at a junction of two uniform tube segments is illustrated in Fig. 2.8a. Note that the transmission coefficients differ from those defined for volume velocity waves (see Fig. 2.7a).

A one-multiplier two-port junction for the sound pressure signals is obtained by rewriting Eqs. (2.32a) and (2.32b) as follows

$$p_k^-(t) = p_{k+1}^-(t) + w(t) \tag{2.33a}$$

$$p_{k+1}^+(t) = p_k^+(t - \tau) + w(t) \tag{2.33b}$$

where

$$w(t) = r_k \Big[p_k^+(t - \tau) - p_{k+1}^-(t) \Big] \tag{2.33c}$$

The one-multiplier scattering junction for sound pressure signals is shown in Fig. 2.8b.

2.2.5 Boundary Conditions of a Cylindrical Tube

Modeling of the end point of a cylindrical tube is achieved with the help of the reflection functions defined above. For a *closed end*, the radiation impedance Z_{rad} tends to infinity. Then the reflection coefficient for the closed end (from the positive side) is

$$r_M = \lim_{Z_{\text{rad}} \to \infty} \frac{Z_{\text{rad}} - Z_M}{Z_{\text{rad}} + Z_M} = 1 \tag{2.34}$$

where Z_M is the characteristic impedance of the last tube section before the closed end. Equation (2.34) means that the wave (either pressure or volume velocity) reflects back from a closed end as such.

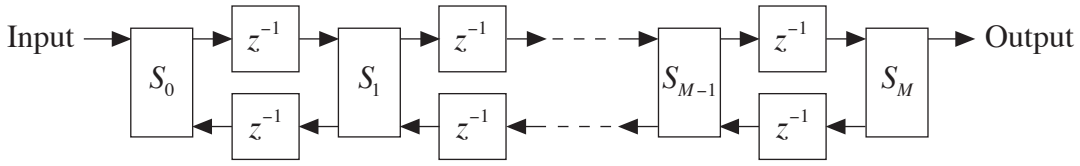

Fig. 2.9 The Kelly–Lochbaum model is a discrete-time implementation of an M-tube model of the vocal tract. The blocks labeled S_n ($n = 0, 1, ..., M$) are two-port scattering junctions.

Modeling of an *open end* of a cylindrical tube is more complicated because the radiation impedance depends on the frequency. Continuous-time models for the radiation impedance $Z_{\text{rad}}(\omega)$ can be found in the acoustics literature. One approximation is the impedance of a vibrating piston at the end of a long tube (see Beranek, 1954, pp. 123–128). The continuous-time radiation impedance can be converted into a discrete-time one using the impulse-invariant transform which is a standard technique in digital signal processing (e.g., Jackson, 1989).

In the case of an open end, a reflection filter, instead of a real reflection coefficient, must be used. It can be expressed in the z-domain as

$$R(z) = \frac{Z_{\text{rad}}(z) - Z_M}{Z_{\text{rad}}(z) + Z_M} \tag{2.35}$$

The radiated volume velocity signal is obtained by filtering the signal at the open end of the tube using the filter $1 - R(z)$ as suggested by Eq. (2.27b) (see also Fig. 2.7). The radiated pressure signal, however, is computed using the filter $1 + R(z)$ (see Fig. 2.8).

In some applications it is of interest to model radiation from an open tube in a flange. This is the case, e.g., in vocal tract modeling. Laine (1982) has derived a simple discrete-time reflection filter that is suitable for speech synthesis. It is based on the use of a first-order differentiator (with a real scaling factor) as a model for the radiation impedance. This approximation leads to a first-order IIR reflection filter. This model has been used successfully as an approximation for the radiation impedance of a woodwind instrument (Välimäki *et al.*, 1992b).

2.2.6 General Properties of the Cylindrical Tube Model

The traditional Kelly–Lochbaum tube model is a ladder or lattice filter with a two-port adaptor between every two unit delays. A model composed of M cylindrical tube sections is illustrated in Fig. 2.9. The model has two delay lines (a digital waveguide) and M+1 two-port scattering junctions S_n.

Bonder (1983a) has derived an expression for the formant frequencies of the cylindrical M-tube model. He has also shown six general properties of the model. These properties are recapitulated below.

1) If the area of each tube is scaled by a common factor, the resonance frequencies remain unchanged. This is a consequence of the fact that the reflection coefficients depend on the ratio of the tube impedances (or areas), not on their exact values [see Eq. (2.26a)].

2) Formant frequencies are inversely proportional to the length of the tube segments. This property suggests that if the lengths of all the tube segments are multiplied

by some factor x, then the resonance frequencies are those of the original model multiplied by $1/x$.

3) A tube model with a certain configuration has the same formant frequencies as the configuration with tubes in the reverse order.

4) In case of an *odd-tube*, i.e., when the number of tube segments is odd (or can be reduced to odd by combining pieces with the same diameter), the model has *fixed formants* at certain frequencies, irrespective of the profile of the tube. These frequencies are given by

$$f_{\mathrm{fix},k} = \left(k + \frac{1}{2}\right)\frac{cM}{2L} \quad \text{for } k = 0,1,2,\ldots \tag{2.36}$$

where c is the speed of sound, M is the total number (odd) of tube sections, and $L = M\ell$ is the overall length of the tube. The indices of the fixed formant frequencies are $[(M+1)/2] + kM$ (for $k = 0,1,2,\ldots$). A model consisting of an even number of sections also has fixed formants if it can be reduced to an odd-tube.

As an example of property 4, let us consider a 3-tube model ($M = 3$) with $L = 17.5$ cm. This model has fixed formants at frequencies 1500 Hz, 4500 Hz, 7500 Hz, etc. These correspond to the resonance frequencies f_2, f_5, f_8, and so on (Bonder, 1983a). The other formants depend on the shape of the tube.

5) The lowest $\lfloor M/2 \rfloor$ formants of an M-tube model determine all the other formants except for the fixed ones: if f is one of the formants of an M-tube model, then also $cM/2L - f$ and $f + kcM/2L$ (for $k = 1,2,3,\ldots$) are its formants. Here $\lfloor \cdot \rfloor$ is the floor function defined by Eq. (3.12) and $cM/2L$ corresponds to the Nyquist frequency f_{N} of the system. Property 5 is another way of saying that the frequency response of a sampled system is periodic with a period of f_{N}.

6) There are $\lfloor (M-1)/2 \rfloor$ different tube configurations with the same formant structure. This feature is a consequence of the previous property. For a 2-tube model there is no degree of freedom because $\lfloor 1/2 \rfloor = 0$. For tube models with $M \geq 3$, there is always a number of different ways to choose the diameters of the sections and still end up with the same formant frequencies. Property 6 is the explanation for the difficulty of estimating the diameters of the tube sections directly from an acoustic signal. This question has been investigated in Bonder (1983b).

2.2.7 Sparse Tube Model

One possible generalization of the M-tube model is such that some of the two-port junctions are removed (or their reflection coefficient is set to 0). This kind of model is more efficient to implement but can still produce essentially a similar spectral envelope as the corresponding acoustic tube system. The number of degrees of freedom is, of course, less than that for the traditional M-tube KL model.

Usually an M-tube model is considered in the context of digital ladder or lattice filters. Probably for this reason, there do not seem to have been many studies of models where fewer than $M + 1$ scattering junctions are employed with M unit delays. Examples are works by Lacroix and Makai (1979), Frank and Lacroix (1986), and Chan

and Leung (1991). From the physical viewpoint this seems like a natural way to reduce the complexity of the model. We shall call a model where less than $M + 1$ junctions are used in connection with M unit delays (with the original sampling frequency) a *sparse tube model*. A further generalization of the M-tube model using fractional delay filtering techniques is introduced in Chapter 4 of this work.

2.2.8 Extensions to the Basic KL Model

There are two severe limitations in the basic Kelly–Lochbaum (KL) model:

1) variable tract length is not incorporated, and

2) all the tube sections have the same length.

Due to the first restriction it is impossible to exactly match given formants, since the vocal tract length is in general not correctly modeled. The second one gives rise to difficulties in the control of the model. Namely, the mapping of the physical parameters to the model parameters (reflection coefficients in the scattering junctions) is not simple. The articulators move in the front-back dimension and the diameter of the tract also varies, whereas in the model only the diameter of the tube can be adjusted at the fixed points. Control would be more intuitive if the junctions could be moved.

There have been two solutions to the these problems. Strube (1975) suggested the use of a *fractional delay element* in place of the last unit delay element of the delay line. He showed that the frequencies and bandwidths of the formants were then changed compared with the ideal ones. Later, Lagrange interpolation was proposed by Laine (1988).

The other method for the accurate adjustment of the vocal tract length is due to Wu *et al*. (1987a, 1987b). They keep the number of sections in the model constant but scale the length of each section by a common factor, which can in principle be an arbitrary real number. This is achieved by changing the sample rate of the model. This, however, does not require that the sample rate of the computer hardware and DA converters should be variable, but the operation is done by software instead. A time-varying FIR interpolating filter is used for sampling-rate conversion. A more efficient form of this technique has been reported by Wright and Owens (1993).

A drawback of both techniques described above is that the length of each tube section cannot be independently controlled in the discrete-time model. In Chapter 4 of this work we present a fractional delay waveguide model where the length of each tube section can be arbitrarily controlled. Consequently, the total length of the vocal tract is continuously variable.

2.3 Junction of Three Acoustic Tubes

A three-port junction of waveguides is needed, for example, in modeling of woodwind instruments. Namely, the *finger holes*, or tone holes, of these musical instruments may be interpreted as short side branches connected to the main bore. Another application is simulation of the human vocal tract with both the oral and the nasal tract models.

In this section, we present the three-port scattering junction. In addition, we study the special case of the three-port junction where two of the three branches have the same impedance, and discuss application of this model to the simulation of finger holes.

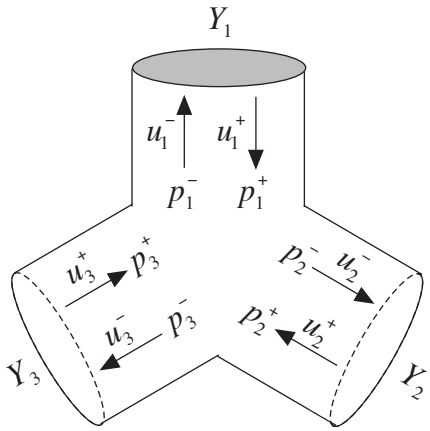

Fig. 2.10 A junction of three cylindrical acoustic tubes.

2.3.1 Three-Port Scattering Junction for Volume Velocity

The equations governing the reflection and transmission of volume velocity waves at a three-port junction are derived first. Generally, for a passive M-way junction, the net volume velocity vanishes, i.e.,

$$\sum_{k=1}^{M} u_k = 0 \tag{2.37}$$

For a three-port junction this yields

$$u_1^+(t) - u_1^-(t) + u_2^+(t) - u_2^-(t) + u_3^+(t) - u_3^-(t) = 0 \tag{2.38}$$

The pressure across the M-way junction must be continuous, that is

$$p_1 = p_2 = \ \ldots \ = p_M \tag{2.39}$$

This rule gives another equation that is needed in the derivation of the three-port junction. Equations (2.37) and (2.39) are analogous to Kirchhoff's current and voltage laws, respectively, used for electrical networks.

In the following the three-port junction for the acoustic *volume velocity* is derived. The variable quantities are illustrated in Fig. 2.10. Superscript '+' means that the signal propagates towards the junction, and '–' denotes the reverse direction. Subscript k refers to the index of the branch.

The pressure in each branch of the junction at point x is expressed in the following form:

$$p_k(x,n) = p_k^+(n-\tau) + p_k^-(n+\tau) = \frac{1}{Y_k}\Big[u_k^+(n-\tau) + u_k^-(n+\tau)\Big] \tag{2.40}$$

where $\tau = x / c$ is the propagation delay and Y_k is the characteristic admittance of the kth branch, defined as $A_k / \rho c$, where A_k is the cross-sectional area of the tube.

According to Eq. (2.39) and (2.40), the pressure signals in the branches of the junction are related in the following way

$$\frac{1}{Y_1}\Big[u_1^+(t)+u_1^-(t)\Big]=\frac{1}{Y_2}\Big[u_2^+(t)+u_2^-(t)\Big]=\frac{1}{Y_3}\Big[u_3^+(t)+u_3^-(t)\Big] \tag{2.41}$$

Using Eqs. (2.38) and (2.41), the outgoing volume velocity signals $u_1^-(t)$, $u_2^-(t)$, and $u_3^-(t)$ may now be solved in terms of the ingoing signals $u_1^+(t)$, $u_2^+(t)$, and $u_3^+(t)$:

$$u_1^-(t)=\frac{(Y_1-Y_2-Y_3)u_1^+(t)+2Y_1\Big[u_2^+(t)+u_3^+(t)\Big]}{Y_1+Y_2+Y_3} \tag{2.42a}$$

$$u_2^-(t)=\frac{(Y_2-Y_1-Y_3)u_2^+(t)+2Y_2\Big[u_1^+(t)+u_3^+(t)\Big]}{Y_1+Y_2+Y_3} \tag{2.42b}$$

$$u_3^-(t)=\frac{(Y_3-Y_1-Y_2)u_3^+(t)+2Y_3\Big[u_1^+(t)+u_2^+(t)\Big]}{Y_1+Y_2+Y_3} \tag{2.42c}$$

Let us define the *reflection coefficient* for the kth branch as

$$r_k=\frac{2Y_k-Y_1-Y_2-Y_3}{Y_1+Y_2+Y_3} \tag{2.43}$$

Using this reflection coefficient for each branch, the Eqs. (2.42) may be rewritten as

$$u_1^-(t)=r_1\,u_1^+(t)+(1+r_1)\Big[u_2^+(t)+u_3^+(t)\Big] \tag{2.44a}$$

$$u_2^-(t)=r_2\,u_2^+(t)+(1+r_2)\Big[u_1^+(t)+u_3^+(t)\Big] \tag{2.44b}$$

$$u_3^-(t)=r_3\,u_3^+(t)+(1+r_3)\Big[u_1^+(t)+u_2^+(t)\Big] \tag{2.44c}$$

2.3.2 Three-Port Scattering Junction for Sound Pressure

We shall next derive the equations governing the scattering of acoustic pressure waves in a three-port junction. The continuity equation (2.38) for volume velocity can be rewritten in terms of pressure as

$$Y_1\Big[p_1^+(t)-p_1^-(t)\Big]+Y_2\Big[p_2^+(t)-p_2^-(t)\Big]+Y_3\Big[p_3^+(t)-p_3^-(t)\Big]=0 \tag{2.45}$$

According to Eq. (2.39) we can write

$$p_1^+(t)+p_1^-(t)=p_2^+(t)+p_2^-(t)=p_3^+(t)+p_3^-(t) \tag{2.46}$$

From the above two equations it is possible to solve for the outgoing pressure signals $p_1^-(t)$, $p_2^-(t)$, and $p_3^-(t)$. This results in the following three equations

$$p_1^-(t)=\frac{(Y_1-Y_2-Y_3)p_1^+(t)+2Y_2p_2^+(t)+2Y_3p_3^+(t)}{Y_1+Y_2+Y_3} \tag{2.47a}$$

$$p_2^-(t) = \frac{(Y_2 - Y_1 - Y_3)p_2^+(t) + 2Y_1 p_1^+(t) + 2Y_3 p_3^+(t)}{Y_1 + Y_2 + Y_3} \tag{2.47b}$$

$$p_3^-(t) = \frac{(Y_3 - Y_1 - Y_2)p_3^+(t) + 2Y_1 p_1^+(t) + 2Y_2 p_2^+(t)}{Y_1 + Y_2 + Y_3} \tag{2.47c}$$

The reflection coefficient for each branch can be defined similarly as earlier for the volume velocity signals. However, the transmission coefficients, which were simply $1 + r_k$ for the kth branch in the case of volume velocity, are now different in every case. Equations (2.47) can be rewritten as

$$p_1^-(t) = r_1\, p_1^+(t) + (1 + r_2)p_2^+(t) + (1 + r_3)p_3^+(t) \tag{2.48a}$$

$$p_2^-(t) = r_2\, p_2^+(t) + (1 + r_1)p_1^+(t) + (1 + r_3)p_3^+(t) \tag{2.48b}$$

$$p_3^-(t) = r_3\, p_3^+(t) + (1 + r_1)p_1^+(t) + (1 + r_2)p_2^+(t) \tag{2.48c}$$

2.3.3 Side Branch in a Uniform Tube

Let us now consider a special case of the three-port junction, namely, a connection of a side branch to a uniform tube. This model is useful in the simulation of a woodwind instrument bore including finger holes (Välimäki *et al.*, 1993b). Figure 2.11 shows the variables used in the following. Note that there are essential differences in the notation with respect to earlier sections. Superscript '+' now means that the signal propagates towards the end (output) of the bore, and '–' means that the signal propagates towards the input. Subscripts 'a' and 'b' refer to the pieces of the bore before and after the side branch connection, respectively, and 's' stands for the side branch.

In this case, two of the three branches have the same characteristic admittance because they are parts of the same uniform tube. We denote the admittances of the uniform tube and the side branch, respectively, by Y_0 and $Y_s(\omega)$. Note that the latter is a function of frequency, and for this reason we use the frequency-domain equivalents of all equations in the following.

The reflection functions for the branches are obtained by substituting $Y_1 = Y_2 = Y_0$ and $Y_3 = Y_s(\omega)$ into Eq. (2.43). This yields

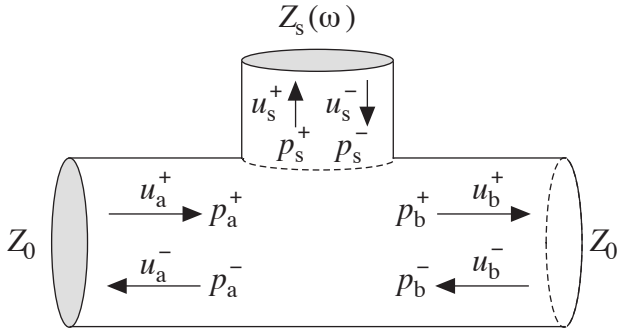

Fig. 2.11 A three-port junction where a side branch is connected to a uniform tube.

$$R_{\mathrm{a}}(\omega) = R_{\mathrm{b}}(\omega) = -\frac{Y_{\mathrm{s}}(\omega)}{Y_{\mathrm{s}}(\omega) + 2Y_0} = -\frac{Z_0}{Z_0 + 2Z_{\mathrm{s}}(\omega)} \tag{2.49}$$

$$R_{\mathrm{s}}(\omega) = \frac{Y_{\mathrm{s}}(\omega) - 2Y_0}{Y_{\mathrm{s}}(\omega) + 2Y_0} = \frac{Z_0 - 2Z_{\mathrm{s}}(\omega)}{Z_0 + 2Z_{\mathrm{s}}(\omega)} \tag{2.50}$$

Note that the configuration in Fig. 2.11 is symmetrical, and hence the reflection coefficients $R_{\mathrm{a}}(\omega)$ and $R_{\mathrm{b}}(\omega)$ are identical and we may define

$$R_0(\omega) = R_{\mathrm{a}}(\omega) = R_{\mathrm{b}}(\omega) \tag{2.51}$$

The reflection function $R_{\mathrm{s}}(\omega)$ can also be written in terms of $R_0(\omega)$:

$$R_{\mathrm{s}}(\omega) = -1 - 2R_0(\omega) \tag{2.52}$$

Side Branch Junction for Volume Velocity

We shall first give the equations for the Fourier transforms of the volume velocity signals $u_{\mathrm{a}}^-(t)$, $u_{\mathrm{b}}^+(t)$, and $u_{\mathrm{s}}^+(t)$ that propagate away from the junction. These formulas are obtained from Eqs. (2.42) by substituting $Y_1 = Y_2 = Y_0$, $Y_3 = Y_{\mathrm{s}}(\omega)$, $p_1^{\pm} = p_{\mathrm{a}}^{\pm}$, $u_2^{\pm} = u_{\mathrm{b}}^{\mp}$, and $u_3^{\pm} = u_{\mathrm{s}}^{\mp}$, and then writing the frequency-domain equivalents of all equations:

$$U_{\mathrm{a}}^-(\omega) = \frac{-Y_{\mathrm{s}}(\omega)U_{\mathrm{a}}^+(\omega) + 2Y_0\left[U_{\mathrm{b}}^-(\omega) + U_{\mathrm{s}}^-(\omega)\right]}{Y_{\mathrm{s}}(\omega) + 2Y_0} \tag{2.53a}$$

$$U_{\mathrm{b}}^+(\omega) = \frac{-Y_{\mathrm{s}}(\omega)U_{\mathrm{b}}^-(\omega) + 2Y_0\left[U_{\mathrm{a}}^+(\omega) + U_{\mathrm{s}}^-(\omega)\right]}{Y_{\mathrm{s}}(\omega) + 2Y_0} \tag{2.53b}$$

$$U_{\mathrm{s}}^+(\omega) = \frac{\left[Y_{\mathrm{s}}(\omega) - 2Y_0\right]U_{\mathrm{s}}^-(\omega) + 2Y_{\mathrm{s}}(\omega)\left[U_{\mathrm{a}}^+(\omega) + U_{\mathrm{b}}^-(\omega)\right]}{Y_{\mathrm{s}}(\omega) + 2Y_0} \tag{2.53c}$$

These equations can be rewritten using the reflection function $R_0(\omega)$ defined by Eq. (2.51) as

$$U_{\mathrm{a}}^-(\omega) = R_0(\omega)U_{\mathrm{a}}^+(\omega) + \left[1 + R_0(\omega)\right]\left[U_{\mathrm{b}}^-(\omega) + U_{\mathrm{s}}^-(\omega)\right] \tag{2.54a}$$

$$U_{\mathrm{b}}^+(\omega) = R_0(\omega)U_{\mathrm{b}}^-(\omega) + \left[1 + R_0(\omega)\right]\left[U_{\mathrm{a}}^+(\omega) + U_{\mathrm{s}}^-(\omega)\right] \tag{2.54b}$$

$$U_{\mathrm{s}}^+(\omega) = -\left[1 + 2R_0(\omega)\right]U_{\mathrm{s}}^-(\omega) - 2R_0(\omega)\left[U_{\mathrm{a}}^+(\omega) + U_{\mathrm{b}}^-(\omega)\right] \tag{2.54c}$$

It is possible to derive an efficient configuration for the side branch junction for volume velocity that is reminiscent of the one-multiplier KL junction developed in Section 2.2.1. This is achieved by regrouping the terms in the above equations. This results in

$$U_{\mathrm{a}}^-(\omega) = U_{\mathrm{b}}^-(\omega) + U_{\mathrm{s}}^-(\omega) + W(\omega) \tag{2.55a}$$

$$U_{\mathrm{b}}^+(\omega) = U_{\mathrm{a}}^+(\omega) + U_{\mathrm{s}}^-(\omega) + W(\omega) \tag{2.55b}$$

$$U_s^+(\omega) = -U_s^-(\omega) - 2W(\omega) \tag{2.55c}$$

where we have defined

$$W(\omega) = R_0(\omega)\left[U_a^+(\omega) + U_b^-(\omega) + U_s^-(\omega)\right] \tag{2.55d}$$

The term $W(\omega)$ is common to Eqs. (2.55a)–(2.55c) and the reflection function $R_0(\omega)$ appears only in Eq. (2.55d). Thus the implementation of this side branch junction involves only one filtering operation.

In Välimäki *et al.* (1993b) this result was derived using characteristic impedances instead of admittances. A special case of the three-port junction, where the volume velocity u_s^- was assumed to be zero, has been presented by Borin *et al.* (1990).

Side Branch Junction for Sound Pressure

We shall now derive the model for the side branch junction in terms of sound pressure signals. The equations for the pressure signals are obtained from Eqs. (2.47) by substituting $Y_1 = Y_2 = Y_0$, $Y_3 = Y_s(\omega)$, $p_1^{\pm} = p_a^{\pm}$, $p_2^{\pm} = p_b^{\mp}$, and $p_3^{\pm} = p_s^{\mp}$. This yields

$$P_a^-(\omega) = \frac{-Y_s(\omega)P_a^+(\omega) + 2Y_0P_b^-(\omega) + 2Y_s(\omega)P_s^-(\omega)}{Y_s(\omega) + 2Y_0} \tag{2.56a}$$

$$P_b^+(\omega) = \frac{-Y_s(\omega)P_b^-(\omega) + 2Y_0P_a^+(\omega) + 2Y_s(\omega)P_s^-(\omega)}{Y_s(\omega) + 2Y_0} \tag{2.56b}$$

$$P_s^+(\omega) = \frac{\left[Y_s(\omega) - 2Y_0\right]P_s^-(\omega) + 2Y_0\left[P_a^+(\omega) + P_b^-(\omega)\right]}{Y_s(\omega) + 2Y_0} \tag{2.56c}$$

These equations can be simplified, using the reflection function $R_0(\omega)$ that has been defined by Eq. (2.51), as

$$P_a^-(\omega) = R_0(\omega)P_a^+(\omega) + \left[1 + R_0(\omega)\right]P_b^-(\omega) - R_0(\omega)P_s^-(\omega) \tag{2.57a}$$

$$P_b^+(\omega) = R_0(\omega)P_b^-(\omega) + \left[1 + R_0(\omega)\right]P_a^+(\omega) - R_0(\omega)P_s^-(\omega) \tag{2.57b}$$

$$P_s^+(\omega) = -\left[1 + 2R_0(\omega)\right]P_s^-(\omega) + \left[1 + R_0(\omega)\right]\left[P_a^+(\omega) + P_b^-(\omega)\right] \tag{2.57c}$$

Unfortunately, it is not possible to derive a form where the filtering by $R_0(\omega)$ needs to be computed only once, as was the case for the volume velocity signals. If, however, the pressure signal p_s^- is zero—which is a reasonable assumption in many practical cases—then an economical implementation can be derived. The result is

$$P_a^-(\omega) = P_b^-(\omega) + W(\omega) \tag{2.58a}$$

$$P_b^+(\omega) = P_a^+(\omega) + W(\omega) \tag{2.58b}$$

$$P_s^+(\omega) = P_a^-(\omega) + P_b^+(\omega) + W(\omega) \tag{2.58c}$$

where we have defined

$$W(\omega) = R_0(\omega)\left[P_a^+(\omega) + P_b^-(\omega)\right] \tag{2.58d}$$

and assumed that $P_s^-(\omega) = 0$.

2.4 Waveguide Modeling of Conical Tubes

In this section we discuss modeling of acoustic tube systems that are constructed of truncated cones. The basic idea of simulating non-cylindrical tubes using digital waveguides was presented by Smith (1991). Formerly, conical tubes have been modeled by approximating the profile of the tube by cylindrical tube sections (see, e.g., Caussé *et al.*, 1984; Cook, 1988). As the number of the cylindrical sections is increased, the approximation approaches the ideal solution. The same idea can as well be applied to other tube profiles, such as exponential or catenoidal horns that are used in brass wind instruments and loudspeakers.

The motivation for the conical tube models is that many practical acoustic systems are constructed of conical sections. For example, most woodwind instruments, e.g., the saxophones, oboes, and bassoons, have a conical bore. The reason for the success of cylindrical and conical bores in musical acoustics is that they are *non-dispersive* waveguides and thus support harmonically related modes (Fletcher and Rossing, 1991).

An important field of application for conical modeling is approximation of other non-cylindrical tube shapes. For these approximations, a model consisting of conical sections will converge more quickly towards the ideal solution than when using the traditional cylindrical sections (Caussé *et al.*, 1984). This approach offers another extension to the modeling of the human vocal tract. Figure 2.12 shows the basic idea of this technique.

In this new kind of articulatory model, the diameter profile of the vocal tract is piecewise linear. This may be interpreted as a *first-order polynomial approximation* of the profile function. With this vocal tract model, synthetic speech signals with realistic

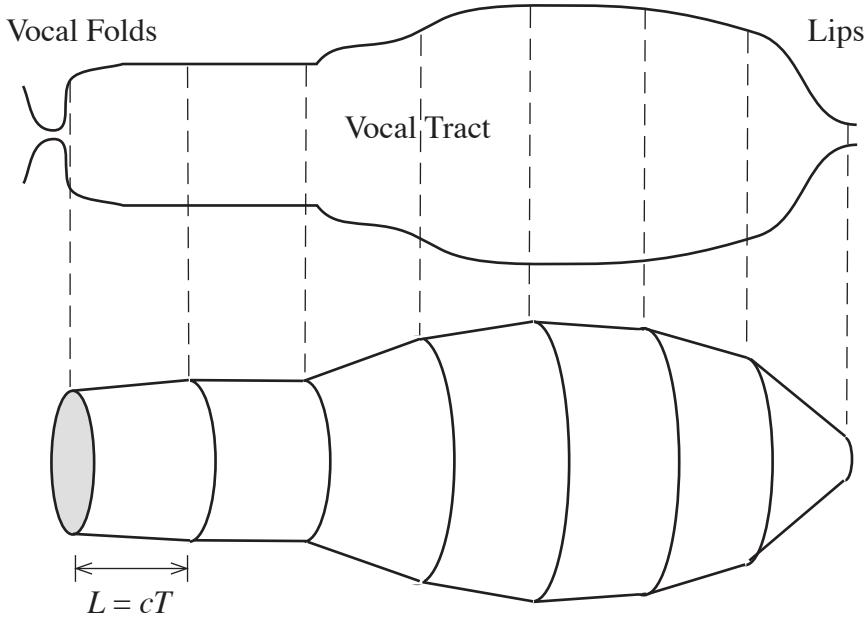

Fig. 2.12 Approximation of the vocal tract profile using conical tube sections.

formant structure can be obtained using a sparse tube model with only a few sections.

In this section, we first introduce a digital filter model for a junction of two conical tube sections with different tapers and diameters. A useful special case is a junction of two conical sections with different tapers but the same diameter. This is applicable to the modeling of at least partly conical acoustic tubes as well as modeling of arbitrary tube profiles. In addition, a discrete-time model for the radiation and reflection at the open end of a conical tube is developed. This model can be applied to the modeling of wind instruments.

We first discuss wave propagation in a conical tube and derive the reflection and transmission functions that govern the scattering of pressure waves in a junction of two conical tube section.

2.4.1 Traveling Wave Solution for Spherical Waves

The wave equation for spherical waves propagating in a conical tube is obtained by writing the general three-dimensional wave equation in spherical coordinates and assuming that the pressure p is a function of r and t only. This yields

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p}{\partial r}\right) = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{2.59}$$

where r is the distance from the cone apex. This may be rewritten as

$$\frac{\partial^2 rp}{\partial t^2} = c^2\frac{\partial^2 rp}{\partial r^2} \tag{2.60}$$

Substituting $\psi = rp$ it is seen that the equation is of the same form as the one-dimensional wave equation [see Eq. (2.1)]

$$\frac{\partial^2 \psi}{\partial t^2} = c^2\frac{\partial^2 \psi}{\partial r^2} \tag{2.61}$$

The general solution thus has a form similar to that given by Eq. (2.2), that is

$$\psi(r,t) = f^+(r - ct) + f^-(r + ct) \tag{2.62}$$

where f^+ and f^- are two traveling waves of arbitrary shape. The solution in terms of pressure p is obtained by substituting $\psi = rp$ so that

$$p(r,t) = \frac{1}{r}f^+(r - ct) + \frac{1}{r}f^-(r + ct) \tag{2.63}$$

where $(1/r)f^+(r - ct)$ represents a pressure wave component propagating in the positive x direction with speed c whereas $(1/r)f^-(r + ct)$ represents a pressure wave propagating in the negative x direction with the same speed.

The solution is similar to Eq. (2.2) except for the factor $1/r$. This difference is interpreted so that the amplitude of a pressure wave is decreased as it travels further away from the apex of the cone and, equivalently, the pressure is increased as the wave travels towards the apex. Note that the amplitude of the wave would be infinite at $r = 0$, and therefore it is not possible to consider the behavior of the wave at the cone apex.

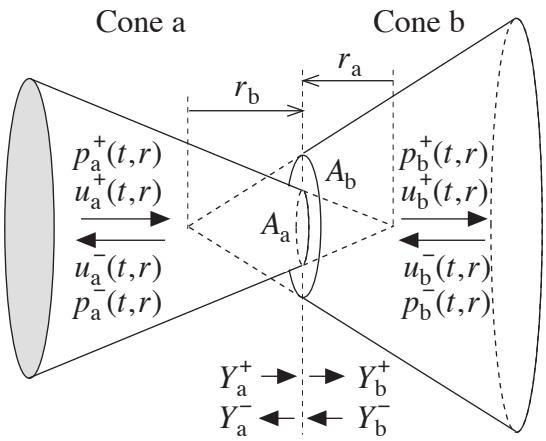

Fig. 2.13 A junction of conical tubes that have a different diameter at the junction. In this example $A_b > A_a$, $r_a < 0$, and $r_b > 0$.

2.4.2 Junction of Conical Tube Sections

In the following we derive equations that govern the reflection and transmission of pressure waves at the junction of two conical tubes of different diameter and taper. We shall not discuss the scattering of volume velocity waves, since this case is essentially more complicated (see Ayers *et al.*, 1985).

The following conditions must be fulfilled at the junction:

$$p_a^+(r,t) + p_a^-(r,t) = p_b^+(r,t) + p_b^-(r,t) \tag{2.64}$$

$$u_a^+(r,t) - u_a^-(r,t) = u_b^+(r,t) - u_b^-(r,t) \tag{2.65}$$

The physical interpretation of the above and the following equations is depicted in Fig. 2.13. Our notation follows the guidelines of Gilbert *et al.* (1990)†.

The admittances in different directions determine the ratio of volume velocity and pressure components of the propagating wave, that is

$$Y_a^{\pm} = \frac{u_a^{\pm}}{p_a^{\pm}} \tag{2.66a}$$

$$Y_b^{\pm} = \frac{u_b^{\pm}}{p_b^{\pm}} \tag{2.66b}$$

On the other hand, the acoustic admittances of spherical waves in a conical tube are defined as (see, e.g., Kinsler and Frey, 1950, pp. 300–303)

$$Y_a^{\pm}(\omega) = \frac{A_a}{\rho c}\left(1 \mp \frac{j}{k r_a}\right) \tag{2.67a}$$

† Note that in some references, e.g., Martínez and Agulló (1988) and Gilbert *et al.* (1990), the superscripts '+' and '–' are used to denote opposite directions with respect to our notation.

$$Y_{\mathrm{b}}^{\pm}(\omega) = \frac{A_{\mathrm{b}}}{\rho c}\left(1 \mp \frac{j}{kr_{\mathrm{b}}}\right) \tag{2.67b}$$

where $k = \omega / c$ is the wave number. Note that the acoustic admittance is always a function of frequency in the case of conical tubes. For this reason we use frequency-domain description in the following.

Based on Eqs. (2.65) and (2.66) we may write

$$P_{\mathrm{a}}^{+}(\omega)Y_{\mathrm{a}}^{+}(\omega) - P_{\mathrm{a}}^{-}(\omega)Y_{\mathrm{a}}^{-}(\omega) = P_{\mathrm{b}}^{+}(\omega)Y_{\mathrm{b}}^{+}(\omega) - P_{\mathrm{b}}^{-}(\omega)Y_{\mathrm{b}}^{-}(\omega) \tag{2.68}$$

Eliminating $P_{\mathrm{b}}^{+}(\omega)$ using Eq. (2.6) and solving for $P_{\mathrm{a}}^{-}(\omega)$ yields

$$P_{\mathrm{a}}^{-}(\omega) = \frac{Y_{\mathrm{a}}^{+}(\omega) - Y_{\mathrm{b}}^{+}(\omega)}{Y_{\mathrm{a}}^{-}(\omega) + Y_{\mathrm{b}}^{+}(\omega)} P_{\mathrm{a}}^{+}(\omega) + \frac{Y_{\mathrm{b}}^{+}(\omega) + Y_{\mathrm{b}}^{-}(\omega)}{Y_{\mathrm{a}}^{-}(\omega) + Y_{\mathrm{b}}^{-}(\omega)} P_{\mathrm{b}}^{-}(\omega) \tag{2.69}$$

Here the first term determines the pressure signal that reflects back from the junction while the second term determines the transmitted pressure signal. Thus we may define the reflection function $R^{+}(\omega)$ and the transmission function $T^{-}(\omega)$ in the following way:

$$R^{+}(\omega) = \frac{Y_{\mathrm{a}}^{+}(\omega) - Y_{\mathrm{b}}^{+}(\omega)}{Y_{\mathrm{a}}^{-}(\omega) + Y_{\mathrm{b}}^{+}(\omega)} \tag{2.70a}$$

$$T^{-}(\omega) = \frac{Y_{\mathrm{b}}^{+}(\omega) + Y_{\mathrm{b}}^{-}(\omega)}{Y_{\mathrm{a}}^{-}(\omega) + Y_{\mathrm{b}}^{+}(\omega)} \tag{2.70b}$$

Similarly, we may eliminate $P_{\mathrm{a}}^{-}(\omega)$ from Eq. (2.68) and solve for $P_{\mathrm{b}}^{+}(\omega)$. We then obtain

$$P_{\mathrm{b}}^{+}(\omega) = \frac{Y_{\mathrm{a}}^{+}(\omega) + Y_{\mathrm{a}}^{-}(\omega)}{Y_{\mathrm{b}}^{+}(\omega) + Y_{\mathrm{a}}^{-}(\omega)} P_{\mathrm{a}}^{+}(\omega) + \frac{Y_{\mathrm{b}}^{-}(\omega) - Y_{\mathrm{a}}^{-}(\omega)}{Y_{\mathrm{b}}^{+}(\omega) + Y_{\mathrm{a}}^{-}(\omega)} P_{\mathrm{b}}^{-}(\omega) \tag{2.71}$$

The reflection and transmission functions are now expressed as

$$R^{-}(\omega) = \frac{Y_{\mathrm{b}}^{-}(\omega) - Y_{\mathrm{a}}^{-}(\omega)}{Y_{\mathrm{b}}^{+}(\omega) + Y_{\mathrm{a}}^{-}(\omega)} \tag{2.72a}$$

$$T^{+}(\omega) = \frac{Y_{\mathrm{a}}^{+}(\omega) + Y_{\mathrm{a}}^{-}(\omega)}{Y_{\mathrm{b}}^{+}(\omega) + Y_{\mathrm{a}}^{-}(\omega)} \tag{2.72b}$$

Now we can derive general formulas for the reflection and transmission functions by substituting Eqs. (2.67a) and (2.67b). After some algebraic manipulation this yields

$$R^{+}(\omega) = \frac{B-1}{B+1} - \frac{2B\alpha}{(B+1)(j\omega + \alpha)} \tag{2.73a}$$

$$R^{-}(\omega) = \frac{1-B}{1+B} - \frac{2\alpha}{(1+B)(j\omega + \alpha)} \tag{2.73b}$$

$$T^+(\omega) = \frac{2B}{B+1}\left(1 - \frac{\alpha}{j\omega + \alpha}\right) = 1 + R^+(\omega) \tag{2.73c}$$

$$T^-(\omega) = \frac{2}{B+1}\left(1 - \frac{\alpha}{j\omega + \alpha}\right) = 1 - R^-(\omega) \tag{2.73d}$$

where B is the area ratio of the conical sections 'a' and 'b' at the junction, i.e.,

$$B = \frac{A_\mathrm{a}}{A_\mathrm{b}} \tag{2.74}$$

and α is defined by

$$\alpha = -\frac{c}{A_\mathrm{a} + A_\mathrm{b}}\left(\frac{A_\mathrm{a}}{r_\mathrm{a}} - \frac{A_\mathrm{b}}{r_\mathrm{b}}\right) \tag{2.75}$$

A major difference between the junction of cylindrical tubes used in the KL model and of the conical tubes presented above is that now the scattering is described by means of a filter instead of a real coefficient. Note, however, that in limit $r_\mathrm{a}, r_\mathrm{b} \to \infty$ the reflection and transmission functions given by Eqs. (2.73) approach the reflection and transmission coefficients for a junction of two cylindrical tubes specified for pressure waves (see Section 2.2.4).

The expression for the reflection function R^+ of a junction of conical tubes of different diameters and tapers has also been derived by Martínez and Agulló (1988) but using another approach. Also, they use the subscripts '+' and '–' to mean exactly the opposite.

Above we have derived equations for the reflection and transmission functions of a junction of two conical tubes using acoustic admittances. This choice was made because it leads to simpler formulas. As an example we give the equation for $R^+(\omega)$ using acoustic impedances to show the difference between these two formulations:

$$R^+(\omega) = \frac{\dfrac{1}{Z_\mathrm{a}^+(\omega)} - \dfrac{1}{Z_\mathrm{b}^+(\omega)}}{\dfrac{1}{Z_\mathrm{a}^-(\omega)} + \dfrac{1}{Z_\mathrm{b}^+(\omega)}} = \frac{Z_\mathrm{b}^+(\omega) - Z_\mathrm{a}^+(\omega)}{\dfrac{Z_\mathrm{a}^+(\omega) Z_\mathrm{b}^+(\omega)}{Z_\mathrm{a}^-(\omega)} + Z_\mathrm{a}^+(\omega)} \tag{2.76}$$

This form should be compared with Eq. (2.70a).

2.4.3 Reflection Function with Taper Discontinuity Only

A useful special case of the junction of two conical tube sections is obtained by setting areas A_a and A_b to be equal (see Fig. 2.14). This implies that $B = 1$. Substituting this into Eqs. (2.73) we obtain the reflection and transmission functions for a junction with taper discontinuity only, that is

$$R^+(\omega) = R^-(\omega) = -\frac{\alpha}{j\omega + \alpha} \tag{2.77a}$$

$$T^+(\omega) = T^-(\omega) = 1 - \frac{\alpha}{j\omega + \alpha} \tag{2.77b}$$

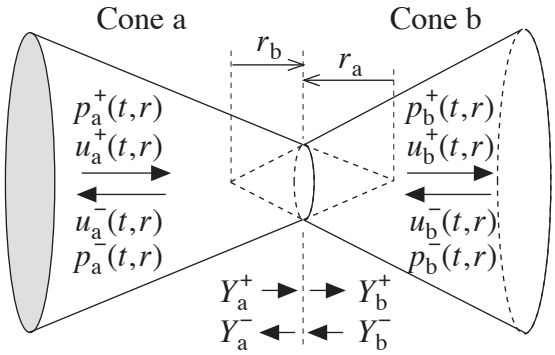

Fig. 2.14 A junction of conical tubes that have the same diameter A at the junction. In this example $r_a < 0$ and $r_b > 0$.

where α is defined as in Eq. (2.75). However, now when the areas A_a and A_b are equal, the coefficient α is simplified into the form

$$\alpha = -\frac{c}{2r_a} + \frac{c}{2r_b} = \frac{c(r_a - r_b)}{2r_a r_b} \tag{2.78}$$

where c is the speed of sound, and r_a and r_b are the distances from the (imagined) tips of cones 'a' and 'b', respectively, as illustrated in Fig. 2.14. Note that now the reflection and transmission functions given by Eqs. (2.77) are the same seen from both sides of the junction.

2.4.4 Closed and Open End of a Truncated Conical Tube

When the end of a truncated conical tube is closed, the radiation impedance $Z_b^+(\omega)$ tends to infinity, which implies that the radiation admittance $Y_b^+(\omega)$ tends to zero. Then the reflection function of a spherical wave is given by

$$R^+(\omega) = \frac{Y_a^+(\omega) - 0}{Y_a^-(\omega) + 0} = \frac{1 - j/kr_a}{1 + j/kr_a} = \frac{j\omega + c/r_a}{j\omega - c/r_a} = 1 + \frac{2c/r_a}{j\omega - c/r_a} \tag{2.79}$$

This equation can be used for modeling the reflection of acoustic wave at the closed end of a conical tube. For $r_a = 0$ Eq. (2.79) is not well defined. This is not a problem in practice because the case is not physically meaningful. Note that in the limit $r_a \to \infty$ the transfer function approaches unity, which is the reflection coefficient for a closed end of a lossless cylindrical tube (see Section 2.2.5).

The acoustic impedance of an open end of a conical tube has not been formulated. Thus it is not possible to derive the corresponding reflection function. Martínez and Agulló (1988) have proposed an *ad hoc* continuous-time reflection function that can be used as a first approximation:

$$r_0(t) = \begin{cases} -a^2 t e^{-at} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{2.80a}$$

with

$$a = c\left(\frac{0.787e}{D_o} - \frac{1}{2r_a}\right) \tag{2.80b}$$

where c is the speed of sound, D_o is the diameter of the open end, r_a the distance of the opening from the apex of the conical tube section, and e the Neperian ($e = 2.71828$).

The function given by Martínez and Agulló approximates the reflection of a pressure wave from the open end of a *conical tube in an infinite wall*. This assumption is often used in models of speech production. The above reflection function is not a very accurate approximation as may be observed by comparing the impulse responses shown in Martínez and Agulló (1988), but it may be good enough for a speech synthesizer.

Note that the case of a conical tube in an infinite wall is not interesting for modeling wind instruments. Rather, we should find a proper approximation for the situation where there is no wall of any kind, but the tube is in free space.

Another approximation for the radiation impedance of the open end of a conical tube is obtained by using the impedance of a spherical source of an equivalent radiation area. Interestingly enough, this impedance has the same form as that of an infinite conical section.

2.4.5 Discrete-Time Reflection and Transmission Filters

In this section we discuss digital waveguide modeling of conical tube systems. We convert the continuous-time reflection functions derived in Section 2.1 into discrete-time filters using the impulse-invariant transformation. This approach has been suggested by Smith (1987) to be used in digital waveguide modeling.

The reflection function for the taper and diameter discontinuity in a conical tube is given by Eq. (2.73a) as

$$R(\omega) = \frac{B-1}{B+1} - \frac{2B\alpha}{(B+1)(j\omega+\alpha)} \tag{2.81}$$

where the coefficient α is defined by Eq. (2.75). The digital filter for the reflection junction is obtained applying the impulse-invariant transform to Eq. (2.81):

$$R(z) = \frac{B-1}{B+1} - \frac{2BT\alpha/(B+1)}{1-e^{-\alpha T}z^{-1}} \tag{2.82}$$

In many applications, such as modeling of wind instruments bores or the human vocal tract, it is needed to model a junction of conical tubes with no diameter discontinuity. Since the reflection filters $R^+(\omega)$ and $R^-(\omega)$ are in this case the same and the transmission filters are defined by $T^+(\omega) = 1 + R^+(\omega)$ and $T^-(\omega) = 1 + R^-(\omega)$, we can use a single reflection filter $R(\omega) = R^+(\omega)$. The impulse-invariant transform of the reflection function given by Eq. (2.77a) yields a first-order all-pole filter

$$R(z) = \frac{b_0}{1 + a_1 z^{-1}} \tag{2.83a}$$

with the filter coefficient

$$a_1 = -e^{-\alpha T} \tag{2.83b}$$

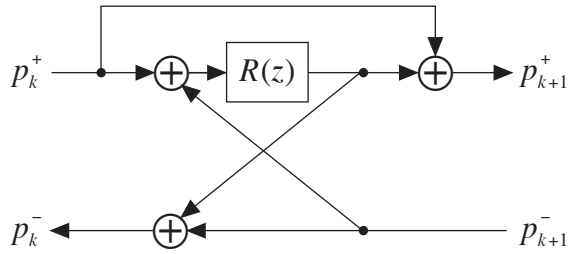

Fig. 2.15 The two-port junction that models the scattering of pressure signals at the junction of two conical tubes when there is no discontinuity of diameter.

where α is defined as in Eq. (2.78). It is necessary to scale this transfer function to have a gain of –1 at $\omega = 0$. This is achieved by setting

$$b_0 = -(1 + a_1) = -1 + e^{-\alpha T} \tag{2.83c}$$

The discrete-time model for the junction of two conical tubes that have the same diameter at the junction is illustrated in Fig. 2.15.

Figure 2.16 shows the magnitude responses of the continuous-time filter $|R(\omega)|$ and its discrete-time version $|R(e^{j\omega})|$ (obtained with the impulse-invariant method) with five

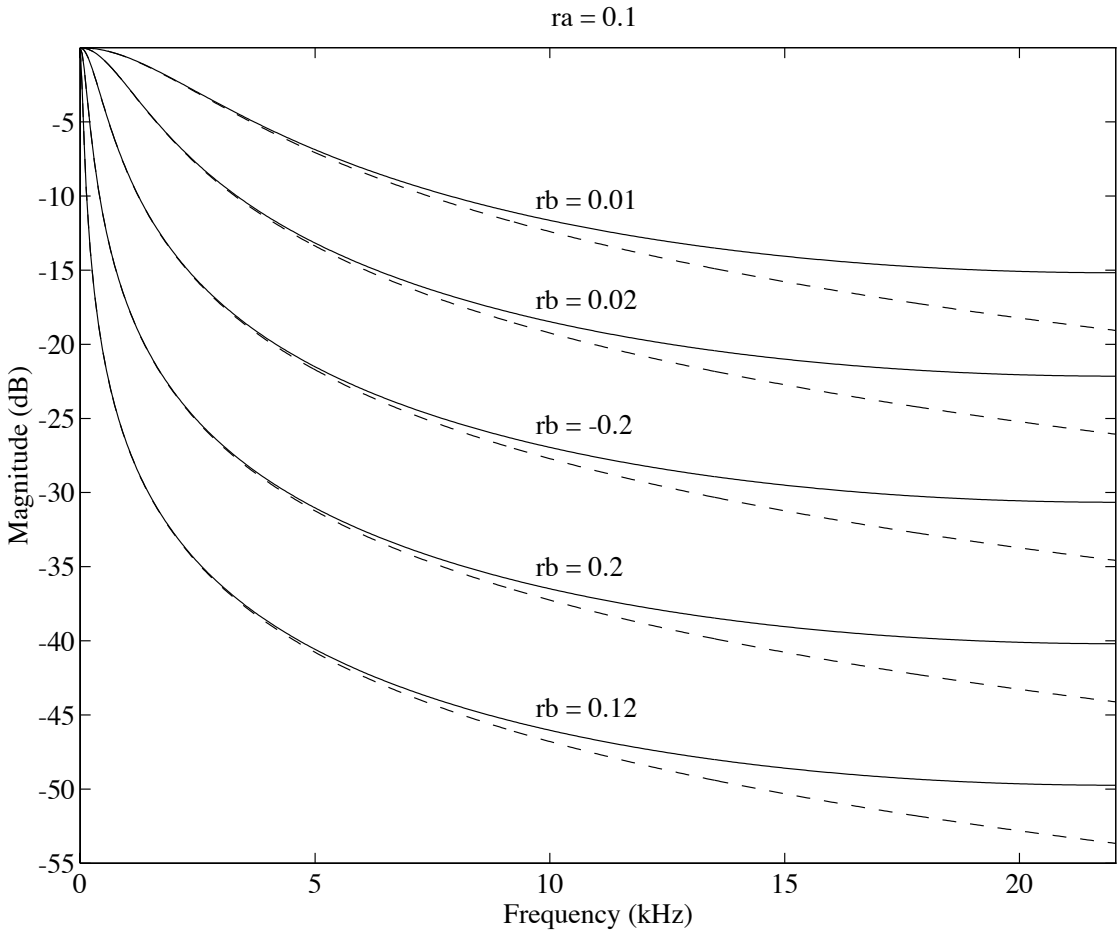

Fig. 2.16 Magnitude responses of the analog (dashed lines) and digital (solid lines) reflection functions with taper discontinuity only. The sampling rate is 44.1 kHz.

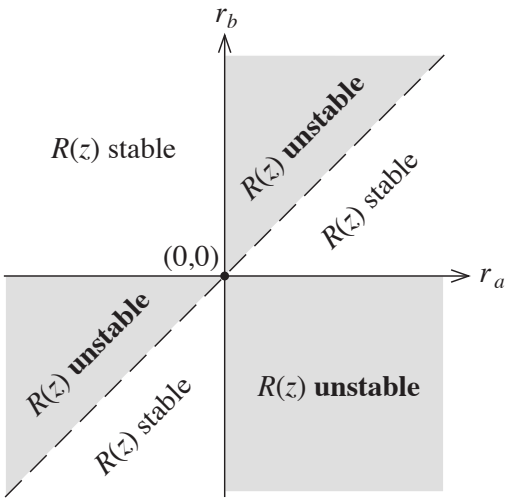

Fig. 2.17 The stability of the reflection filter $R(z)$ depends on the parameters r_a and r_b. The filter is unstable in the shaded regions and stable in the white ones.

different values of r_b. The parameter r_a is equal to 0.1 m in these examples. The effect of aliasing due to the impulse-invariant transformation is clearly seen by comparing the curves of the analog (dotted lines) and the digital (solid lines) filters. Note, however, that the difference between the magnitude response pairs is several dB only at frequencies over 10 kHz. At low frequencies, which are of great importance, the digital approximations coincide with the analog ones.

2.4.6 Stability of the Reflection Filter

The reflection filter $R(z)$ defined by Eq. (2.83) is stable when $e^{-\alpha T} < 1$ which implies that α must be positive (since $T > 0$). The sign of α depends on the values of r_a and r_b according to Eq. (2.78).

The regions of stability and instability of $R(z)$ are illustrated in Fig. 2.17 on a parameter plane where the abscissa corresponds to the value of r_a and the ordinate to the value of r_b. There are three distinct regions where the filter $R(z)$ is stable and three where it is unstable. The unstable regions include half of all possible conical tube configurations. These cases may be impossible to avoid in practice.

Fortunately it appears that having an unstable filter as a part of a larger system does not necessarily imply that the overall system is unstable. Physical systems are passive and they are always stable. Also digital models that simulate physically realizable systems have to be stable.

2.4.7 Modeling the Ends of a Conical Tube

The reflection of pressure waves at an open or a closed end of a conical tube depends on the frequency as was discussed in Section 2.4.4. The continuous-time formulas that describe the reflection at the end of a conical tube may be converted into digital filters using the impulse-invariant transformation.

The impulse-invariant transform of the reflection function (2.79) for the closed end of a conical tube yields a first-order IIR filter

$$R(z) = 1 + \frac{b_0}{1 + a_1 z^{-1}} \tag{2.84}$$

with the coefficients $b_0 = 2cT / r_a$ and $a_1 = -e^{cT/r_a}$. This transfer function is stable when $r_a < 0$ but unstable when $r_a > 0$.

The open end of a conical tube in an infinite wall can be modeled using the continuous-time reflection function of Eq. (2.80) designed by Martínez (Martínez and Agulló, 1988). It is easily transformed into the frequency domain by using the property that differentiation in the frequency domain corresponds to multiplication by time in the time domain. Thus, the Laplace transform may be written as

$$R_0(s) = -a^2 \frac{d}{ds}\left(\frac{1}{s + a}\right) = -\frac{a^3}{(s + a)^2} \tag{2.85}$$

The corresponding z-transform obtained by the impulse-invariant method is given by

$$R_0(z) = -a^2 \frac{d}{dz}\left(\frac{1}{1 - e^{-aT} z^{-1}}\right) \tag{2.86}$$

This yields

$$R_0(z) = \frac{b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{2.87a}$$

with the coefficients

$$b_1 = -a^2 T e^{-aT}, \quad a_1 = -2e^{-aT}, \quad a_2 = e^{-2aT} \tag{2.87b}$$

The impulse response of the discrete-time filter given by Eq. (2.87) is a sampled version of Eq. (2.80a), that is

$$h(nT) = -a^2 nT e^{-anT} \quad \text{for } n = 0, 1, 2, \ldots \tag{2.88}$$

2.4.8 Conclusions and Discussion

This section has discussed computational modeling of conical acoustic tubes. A tube system consisting of truncated cones was studied. An interesting special case of this configuration is one with no discontinuity in the diameter between the contiguous conical tube sections, that is, only the taper of the tube is changed at the junction. This kind of model may be utilized for conical waveguide simulation of the human vocal tract or the bores of wind instruments.

The waveguide simulation of acoustic tubes constructed of conical parts is a new area of research. It was first discussed by Välimäki and Karjalainen (1994a, 1994b), see also Välimäki (1994b). Recently, de Bruin and Walstijn (1995) have implemented a conical waveguide model following the same guidelines as presented here (see also Walstijn and de Bruin, 1995).

Several questions are left for future research. For example, waveguide simulation of finger holes in a conical bore should be studied using the formulations given in Martínez *et al.* (1988). This would be useful for the modeling of woodwind instruments

such as the saxophone. Furthermore, the stability of the overall waveguide model is an issue that should be demonstrated in the general case or at least for certain useful configurations.

Another important topic for future research is to verify the accuracy of the digital model by acoustic measurements. Tube systems including conical parts can be excited by a short pulse and the response measured by a microphone. The actual impulse response is then obtained by deconvolution of the pulse and the response. These results may be compared to those obtained using a waveguide model. Agulló *et al.* (1994, 1995) have measured conical tube systems and showed that the results are in good agreement with the theoretical formulations reported in Martínez and Agulló (1988).

A major unsolved problem is the reflection function for the open end of a conical tube. The exact mathematical solution is obviously involved. Hence we suggest a series of measurements of impulse responses and reflection functions of open conical horns of different taper and diameter. A useful digital reflection filter may then be designed by applying one of the standard techniques used in digital signal processing.

2.5 Waveguide Models for Wind Instruments

One of the first applications of the waveguide modeling technique was a model-based sound synthesis of the clarinet (Smith, 1986; Hirschman, 1991; Välimäki *et al.*, 1992b; Rocchesso and Turra, 1993). A waveguide flute model was introduced in Karjalainen *et al.* (1991) and Välimäki *et al.* (1992a). The block diagram of a general wind instrument model is depicted in Fig. 2.18.

Now the waveguide represents the propagation of wave components in opposite directions in an acoustic tube. The length L of the delay lines is related to the effective length ℓ_{eff} of the tube in the following way:

$$L = \frac{\ell_{\text{eff}} f_{\text{s}}}{c} \tag{2.89}$$

where f_{s} is the sampling rate. Note that in general L is not an integer and again a fractional delay filter must be used to implement a real-valued delay.

The reflection model consists of a digital filter that brings about the frequency-dependent reflection and radiation of the sound wave at the end of the bore. Note that the reflection model has two outputs, one for the outgoing signal and the other for the reflected, ingoing signal.

The excitation model includes a *nonlinearity* which is an essential part of the sound production mechanism in wind instruments. In the case of the clarinet, this system models the operation of a reed that controls the air flow through the mouth piece. The same nonlinearity is also suitable for other reed wind instruments, such as the saxo-

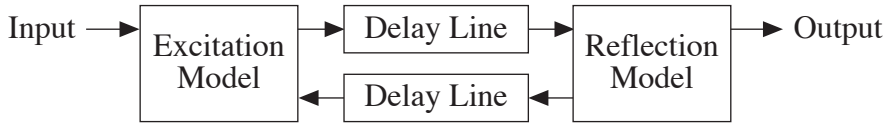

Fig. 2.18 A generic waveguide wind instrument model.

phone (Cook, 1988). In the flute, the nonlinearity models the airflow into the bore and is similar to a sigmoid function (Karjalainen *et al.*, 1991; Välimäki *et al.*, 1992a; Välimäki, 1992).

Cook (1991, 1992) introduced a waveguide model that is capable of synthesizing brass instrument tones. His system is also based on the principle shown in Fig. 2.18. However, the nonlinearity has been obtained by modeling the lips of the player as a mass-spring oscillator. For discrete-time simulation the differential equations have been replaced with difference equations.

2.5.1 Need for Finger-Hole Models

Later in this work we will address the problem of simulating *finger holes* in the bore of a woodwind instrument. Finger holes (or tone holes, or side holes) are a unique feature of woodwinds. They have two essential functions:

1) controlling the fundamental frequency, and

2) influencing the sound quality.

The effect of open finger holes on the sound is characterized by the *open-hole lattice cutoff frequency* (Benade, 1960, 1976). Above this frequency the sound waves mainly radiate out of the tube and produce only weak resonances in the tube. Furthermore, even closed tone holes have their effect, as they effectively lengthen and enlarge the bore in their vicinity (Benade, 1960, 1976).

So far the tone holes have usually been neglected in digital waveguide models for woodwind instruments (see Smith, 1986; Välimäki *et al.*, 1992a, 1992b; Cook, 1992). The obvious limitation in these models is that the fundamental frequency of a synthesized tone can only be controlled by changing the length of the delay lines that form the digital waveguide model for the bore. From the viewpoint of the sound wave propagating inside the bore, this is a very unnatural procedure. Dropping or adding unit delays causes a discontinuity in the signal which can be heard as an annoying click. Another deficiency of simplified woodwind bore modeling is that the tone hole lattice cutoff effect is not automatically included in the model. Thus, its effect has to be added separately, usually during the design of the radiation filter.

In addition to the above mentioned problems, the lack of tone holes limits the playing styles that are available for the user of the woodwind instrument model. *Multiphonics*, for example, cannot be produced by a model without finger holes. When proper finger-hole models are included, multiphonics as well as other special effects, like the key claps, become available to the user of the computer model such as they are for the player of an acoustic instrument.

2.5.2 Modeling of a Single Woodwind Finger Hole

In traditional transmission line models for woodwind instruments a side hole is presented as a T network (Keefe, 1982). A two-by-two linear transformation will convert T parameters to reflections and transmission coefficients. The T network model includes both shunt and series impedances. In practice, the series impedance has a small value and can be neglected (Coltman, 1979).

It is most important to simulate accurately the open finger holes since they determine the fundamental frequency of the tone. Closed holes have a considerably less significant

effect and for simplification they are omitted in our model. In the following a computational model for an open hole is derived.

Model for an Open Finger Hole

The simplest approximation for the acoustic impedance Z_s of a small open hole is pure *acoustic inertance*, that is (see, e.g., Fletcher and Rossing, 1991)

$$Z_s(\omega) = j\omega \frac{\rho \ell}{A_s} \tag{2.90}$$

where ρ is the density of air, A_s is the cross-sectional area of the finger hole, and ℓ is its effective height (see Keefe, 1982, 1990). Multiplication by $j\omega$ in the frequency domain corresponds to differentiation in the time domain. In the discrete-time model the impedance of an open hole can be approximated by means of a *digital differentiator* $H_D(z)$. The transfer function of a digital filter approximating Eq. (2.89) is then

$$\not{Z}_s(z) = \frac{\rho \ell}{A_s} H_D(z) \tag{2.91}$$

Let us denote the discrete-time version of the reflection function of Eq. (2.49) by $R(z)$. This reflection function is expressed as

$$R(z) = -\frac{\not{Y}_s(z)}{\not{Y}_s(z) + 2Y_0} = -\frac{Z_0}{Z_0 + 2\not{Z}_s(z)} \tag{2.92}$$

where $\not{Y}_s(z)$ and $\not{Z}_s(z)$ denote the discrete-time approximations for the continuous-time functions $Y_s(\omega)$ and $Z_s(\omega)$, respectively. When Eq. (2.91) is substituted into Eq. (2.92) the digital reflection filter associated with the finger-hole junction is given by

$$R(z) = -\frac{Z_0}{Z_0 + 2\not{Z}_s(z)} = -\frac{\rho c / A_0}{(\rho c / A_0) + (2\rho \ell / A_s) H_D(z)} \tag{2.93}$$

where A_0 is the cross-sectional area of the bore.

The simplest transfer function of a digital filter approximating the differentiator is

$$H_D(z) = \frac{1}{T}(1 - z^{-1}) \tag{2.94}$$

where T is the sampling interval. This filter computes the difference of two successive input samples and acts as a high-pass filter. Using this approximation Eq. (2.93) can be simplified into the form

$$R(z) = -\frac{1 + a}{1 + az^{-1}} \tag{2.95a}$$

where the filter coefficient is given by

$$a = -\frac{2A_0 \ell}{2A_0 \ell + A_s Tc} \tag{2.95b}$$

Equation (2.95a) is the transfer function of a first-order all-pole filter. Since the coeffi-

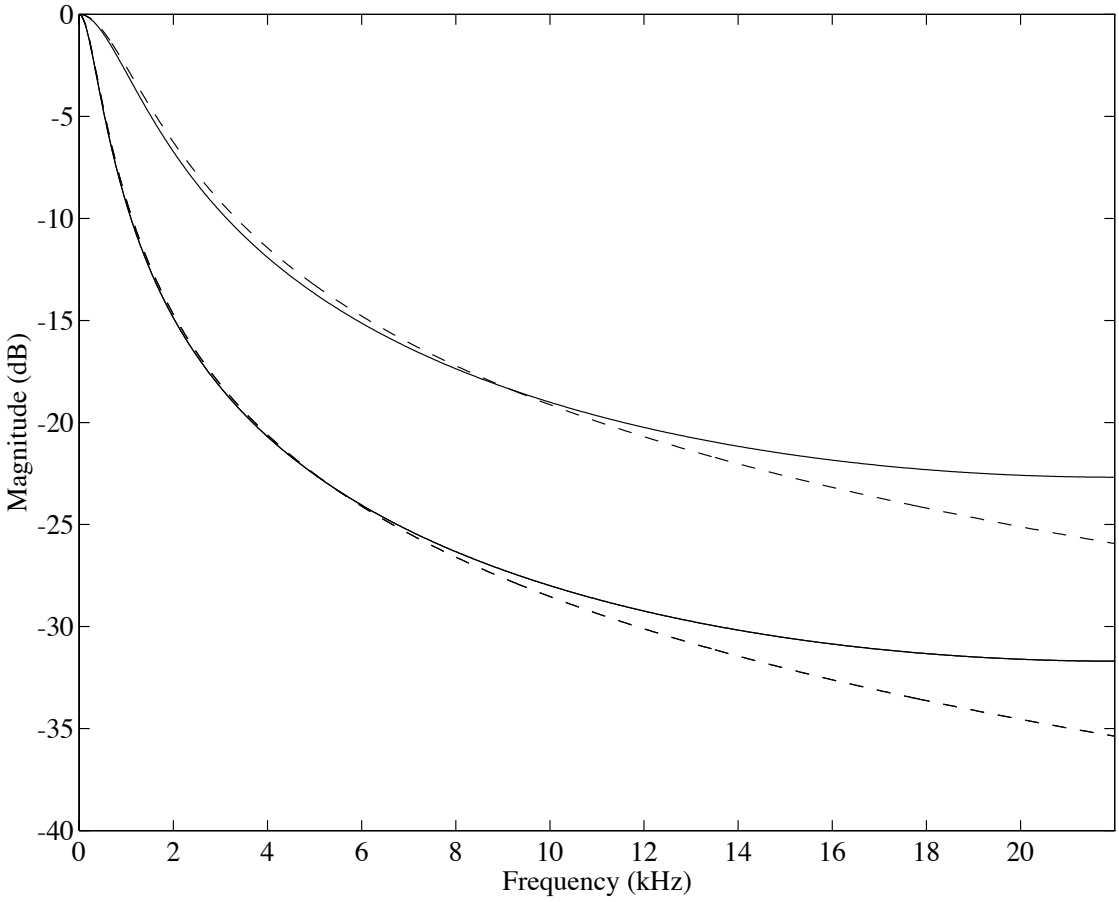

Fig. 2.19 Analog (dashed lines) and digital (solid lines) approximations to the reflection function of a finger-hole junction.

cient a is always negative this is a lowpass filter. At very low frequencies it acts like a constant multiplier, multiplying by –1.

The magnitude responses of the analog reflection function $R_0(\omega)$ computed using Eq. (2.49) with the impedance defined in Eq. (2.90) and the digital reflection function $R(e^{j\omega})$ of Eq. (2.95) are illustrated in Fig. 2.19. The sampling rate used in the discrete-time simulation is 44.1 kHz, i.e., $T = 22.7$ μs. The upper curves are computed for a hole radius of 8.0 mm and an effective height of 17.5 mm, and the lower curves for a hole radius of 6.25 mm and an effective height of 31.7 mm. These examples correspond to two different finger holes of the flute (Coltman, 1979). The bore radius in the flute is 9.5 mm. The approximation error in the magnitude response of the digital filter is less than 1 dB at frequencies below 10 kHz.

Each finger-hole model requires one filter consistent with Eq. (2.95). By controlling the filter coefficient a in the range $[a_0, 0]$ (where a_0 is the nominal value for the open hole) in a suitable way, the finger hole can be closed and opened as in real woodwind instruments to attain desired changes in the effective length of the bore.

2.5.3 Discrete-Time Scattering Junction

Let us consider the implementation of a finger-hole model via a three-port junction as discussed in Section 2.3.3. The frequency-domain Eqs. (2.55) that describe the three-port junction for volume velocity waves can be expressed in the (discrete) time domain

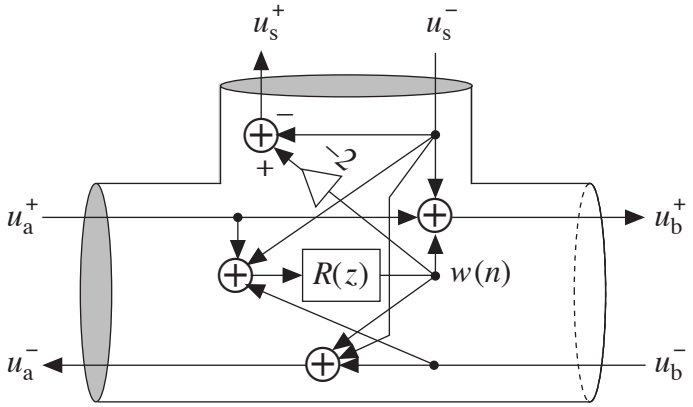

Fig. 2.20 A computational model for scattering of volume velocity waves in a finger-hole junction.

as

$$u_a^-(n-D) = u_b^-(n-D) + u_s^-(n) + w(n) \tag{2.96a}$$

$$u_b^+(n-D) = u_a^+(n-D) + u_s^-(n) + w(n) \tag{2.96b}$$

$$u_s^+(n) = -u_s^-(n) - 2w(n) \tag{2.96c}$$

where D is the real-valued delay corresponding to the position of the center of the finger hole and $w(n)$ is the output of the reflection filter $R(z)$ (see Fig. 2.20), i.e.,

$$w(n) = r(n) * w_1(n) \tag{2.97}$$

where $r(n)$ is the impulse response of the reflection filter $R(z)$, $*$ denotes the convolution sum, and the input signal $w_1(n)$ is

$$w_1(n) = u_a^+(n-D) + u_b^-(n-D) + u_s^-(n) \tag{2.98}$$

The signal flow graph for a finger-hole model is depicted in Fig. 2.20. In this figure the variable quantities are the same as in Fig. 2.11.

2.6 Relation to Other Methods

Smith (1987) has found that there are other approaches that have a lot in common with digital waveguide filters (WGF). Digital *ladder and lattice filters* investigated by Gray and Markel (1973, 1975) are discrete-time structures similar to the WGFs. The normalized ladder filter structure is a special case of a WGF. The KL vocal tract model is also a kind of digital lattice filter.

Digital waveguide filters are also closely related to *wave digital filters* (WDF). In WDFs, circuits are constructed utilizing adaptors that interconnect ports having different impedances (Fettweis, 1972). Adaptors are equivalent to junctions of waveguides. A comprehensive tutorial on WDFs has been written by Fettweis (1986).

2.6.1 Comparison of Waveguide Filters and Wave Digital Filters

Smith (1987, pp. 92–93) has stated the differences between WDFs and WGFs:

- WDFs typically digitize lumped analog filter structures, in general RLC networks, whereas WGFs digitize distributed physical systems such as acoustic resonators.
- In WDFs the frequency variable is transformed into the discrete-time domain via the bilinear transform. In WGFs, the frequency is digitized using the impulse-invariant method.
- In WDFs, the wave variable is a combination of voltage and current. The wave variable in WGFs is typically a basic physical quantity, e.g., the acoustic pressure or volume velocity.

A major motivation for using WDFs in an application is their low sensitivity to changes in multiplier coefficients. Namely, the coefficients of WDFs can be represented with the accuracy of a few bits without changing the behavior of the filter considerably. On the other hand, WGFs are used because they provide a simple, intuitive representation of a physical system. Furthermore, WGFs can be implemented efficiently using a computer.

The Kelly–Lochbaum speech synthesis model is based on a waveguide model for the human vocal tract, although in some references it is stated to be an example of the wave digital filter approach (see, e.g., Schroeter and Sondhi, 1994). This is a reflection of the fact that WDFs and WGFs are closely related. Today WGFs are considered to be a special case of WDFs[†].

Smith (1987, p. 115) criticizes the use of the bilinear transform in WDFs by pointing out that the model is exact only at three frequencies: 0, ∞, and an arbitrary mapping frequency. The waveguide model, on the other hand, utilizes the impulse-invariant method for transforming the frequency variable. Thus wave propagation is modeled accurately at discrete points. Smith states that for this reason it is possible to use voltage or current as a signal variable without having realization problems. In WDFs it is necessary to use a wave variable that is a combination of voltage and current (see, e.g., Fettweis, 1986).

2.6.2 Reverberation and Simulation of Room Acoustics

Reverberators are used to add spatial impression to studio recordings of speech and music and also to change the acoustics of listening rooms. Recursive digital filters including delay lines have been used for implementing artificial reverberators for about 30 years. The initial work in this field is due to Schroeder (1962). A comprehensive study on digital filter structures for reverberation was carried out by Moorer (1979). These techniques have been extended to include a set of delay lines with a feedback matrix (Jot and Chaigne, 1991; Rocchesso, 1993; Rocchesso and Smith, 1994). This kind of system is called a feedback delay network.

Artificial reverberation was the first problem where waveguide filters were applied (Smith, 1985). In this paper Smith defined the digital waveguide as a *lossless bidirectional signal branch*, the simplest special case being a *bidirectional delay line*. As a generalization he mentions a cascade of allpass filters. As we have seen in this chapter,

[†] J. O. Smith, personal communication, November 27, 1995.

a practical digital waveguide system also incorporates losses in terms of gain factors less than one or digital (lowpass) filters whose magnitude response is not greater than unity at any frequency.

The relationship of the feedback delay network and the digital waveguide network has been studied by Smith and Rocchesso (1994). Recently, the digital waveguide technique has been generalized to two dimensions (Van Duyne and Smith, 1993). Further generalization to three or more dimensions is straightforward. Savioja *et al.* (1994) have demonstrated that a three-dimensional waveguide mesh can be used for the simulation of room acoustics.

2.7 Conclusions

In this chapter the fundamentals of waveguide modeling were discussed. Digital waveguides provide a systematic method for designing computational models for physical systems. So far this approach has primarily been applied to one-dimensional resonators since two and three-dimensional extensions have been introduced only recently.

The digital waveguide methods have been used for synthesis of music and speech. Other techniques that lead to similar structures as WGFs were reviewed. These include the Karplus–Strong plucked string algorithm, the Kelly–Lochbaum vocal tract model, delay networks for reverberation, and wave digital filters.

One of the main drawbacks of the WGFs is the discretization of the time variable, which leads to non-accurate modeling of propagation delays. This is the motivation for the use of fractional delays in waveguide models. In the next chapter we discuss the theory of fractional delay filters.